Non-Ideal Epistemic Spaces

JENS CHRISTIAN BJERRING

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
OF THE AUSTRALIAN NATIONAL UNIVERSITY
FEBRUARY 2010

Declaration

This dissertation is soley the work of its author. No part of it has been submitted for any degree or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this dissertation, and all sources used, have been acknowledged.

Acknowledgments

Thanks to Istvan Aranyosi, Ben Blumson, Berit Brogaard, Darren Bradley, Jacek Brzozowski, Fabrizio Cariani, John Cusbert, Kenny Easwaran, Andy Egan, Lina Eriksson, Bill Fish, Johan Gersel, Hilary Greaves, Patrick Greenough, Lloyd Humberstone, Magdalena Balcerak Jackson, Mark Jago, Ole Koksvik, Barak Krakauer, Holly Lawford-Smith, Leon Leontyev, Christian List, Aidan Lyon, Fiona Macpherson, Dan Marshall, John Matthewson, Daniel Nolan, Brian Rabern, Kelly Roe, Susanna Schellenberg, Mike Titlebaum, Clas Weber, and Tobias Wilsch for discussions and comments, provided either in person, in writing, or during talks and seminars. (My apologies to anyone I have accidentally left out in this list.)

Many thanks to Alan Hájek, Brendan Balcerak Jackson, Stephan Leuenberger, Joe Salerno, and Jonathan Schaffer for very useful comments on drafts of various chapters.

But more than anything, I am indebted to David Chalmers, Wolfgang Schwarz, and Weng Hong Tang for invaluable discussion of basically every sentence and thought in this dissertation.

Finally, my deepest thanks to my parents, my brothers, Magnus Agervold, and Leone Miller for support, thoughts and sanity.

Abstract

In a possible world framework, an agent can be said to know a proposition just in case the proposition is true at all worlds that are *epistemically possible* for the agent. Roughly, a world is epistemically possible for an agent just in case the world is not ruled out by anything the agent knows. If a proposition is true at some epistemically possible world for an agent, the proposition is epistemically possible for the agent. If a proposition is true at all epistemically possible worlds for an agent, the proposition is epistemically necessary for the agent, and as such, the agent knows the proposition.

This framework presupposes an underlying space of worlds that we can call *epistemic space*. Traditionally, worlds in epistemic space are identified with *possible worlds*, where possible worlds are the kinds of entities that at least verify all logical truths. If so, given that epistemic space consists solely of possible worlds, it follows that any world that may remain epistemically possible for an agent verifies all logical truths. As a result, all logical truths are epistemically necessary for any agent, and the corresponding framework only allows us to model logically omniscient agents. This is a well-known consequence of the standard possible world framework, and it is generally taken to imply that the framework cannot be used to model non-ideal agents that fall short of logical omniscience.

A familiar attempt to model non-ideal agents within a broadly world involving framework 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 epistemic space, it is easy to avoid logical omniscience. If any logical falsehood is true at some impossible world, then any logical falsehood may remain epistemically possible for some agent. As a result, we can use an impossible world involving framework to model extremely non-ideal agents that do not know *any* logical truths.

A much harder, and considerably less investigated challenge is to ensure that the resulting epistemic 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 impossible worlds that verify complex logical falsehoods, they are nevertheless able to rule out impossible worlds that verify obvious logical falsehoods. To model such agents, we need a construction of a *non-trivial epistemic space* that partly consists of impossible worlds where not "anything goes". This involves imposing substantive constraints on impossible worlds to eliminate from epistemic space, say, trivially impossible worlds that verify obvious logical falsehoods.

The central aim of this dissertation is to investigate the nature of such non-trivially impossible worlds and the corresponding epistemic spaces. To flag my conclusions, I argue that successful constructions of epistemic spaces that can safely navigate between the Charybdis of logical omniscience and the Scylla of "anything goes" are hard, if not impossible to find.

Contents

Chapter 1

Introduction

In a possible world framework, an agent a can be said to know a proposition p just in case p is true at all worlds that are *epistemically possible* for a. Roughly, a world is epistemically possible for a just in case the world is not ruled out by anything a knows. If p is true at some epistemically possible world for a, then p is epistemically possible for a. If p is true at all epistemically possible worlds for a, then p is epistemically necessary for a, and as such, a knows p.

This framework presupposes an underlying space of worlds that we can call *epistemic space*. Traditionally, worlds in epistemic space are identified with *possible worlds*, where possible worlds are the kinds of entities that at least verify all logical truths. If so, given that epistemic space consists solely of possible worlds, it follows that any world that may remain epistemically possible for an agent verifies all logical truths. As a result, all logical truths are epistemically necessary for any agent, and the corresponding framework only allows us to model logically omniscient agents. This is a well-known consequence of the standard possible world framework, and it is generally taken to imply that the framework cannot be used to model non-ideal agents that fall short of logical omniscience.

A familiar attempt to model non-ideal agents within a broadly world involving framework 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 epistemic space, it is easy to avoid logical omniscience. If any logical falsehood is true at some impossible world, then any logical falsehood may remain epistemically possible for some agent. As a result, we can use an impossible world involving framework to model extremely non-ideal agents that do not know *any* logical truths.

A much harder, and considerably less investigated challenge is to ensure that the resulting epistemic 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 impossible worlds that verify complex logical falsehoods, they are nevertheless able to rule out impossible worlds that verify obvious logical falsehoods. To model such agents, we need a construction of a *non-trivial epistemic space* that partly consists of impossible worlds where not "anything goes". This involves imposing substantive constraints on impossible worlds to eliminate from epistemic space, say, trivially impossible worlds that verify obvious logical falsehoods.

The central aim of this dissertation is to investigate the nature of such non-trivially impossible worlds and the corresponding epistemic spaces. To flag my conclusions, I argue that successful constructions of epistemic spaces that can safely navigate between the Charybdis of logical omniscience and the Scylla of "anything goes" are hard, if not impossible to find.[1]

In this chapter, I fill in the gaps that this very rough characterization of the project leaves open and introduce the general framework that I will work within.

[1] The locution of an "anything goes" world and the Charybdis and Scylla formulations are taken from Chalmers (forthcoming).

1.1 The Intuitive Picture[2]

Among the many ways things might be, I know how some of them are. I know that my bike is red and that the sun shines in Canberra. But I do not know whether there is extraterrestrial life. For all I know, there might be life in outer space, and there might not. Nor do I know whether Goldbach's Conjecture is true. For all I know, the conjecture might be true, and it might be false.

Among the many ways things might be, I know a priori how some of them are. I know a priori that my bike is not red and blue all over, and I know a priori that the sun does not both shine and not shine in Canberra. But I do not know a priori whether water is H_2O. For all I know a priori, water might be H_2O, and it might not. Nor do I know a priori whether Fermat's Last Theorem is true. For all I know a priori, the conjecture might be true, and it might be false.

When p might be the case for all an agent knows, we can say that p is *epistemically possible* for the agent. So it is epistemically possible for me that there is extraterrestrial life and that Goldbach's Conjecture is false. Though I have excellent empirical reasons to believe that water is H_2O and that Fermat's Last Theorem is true, I still cannot justify these beliefs using solely a priori reasoning. Intuitively, if I were to suspend all my empirical beliefs and consider whether water is H_2O or whether Fermat's Last Theorem is true, I would not reach a verdict. In this more demanding sense of epistemic possibility, it remains epistemically possible for me that water is not H_2O and that Fermat's Last Theorem is false. Another way to see this is to consider a high stakes bet. Suppose you ask me to bet my laptop with all its contents on the truth of Fermat's Last Theorem for a penny in return. I would decline. When the

[2] The intuitive picture of epistemic possibility presented in this section is heavily influenced by Chalmers (forthcoming): pp. 1-2.

stakes are high, it is epistemically possible for me that Fermat's Last Theorem is false and that water is not H_2O. In contrast, offer me any bet that pays a penny if it is true that my bike is not red and blue over all, and I would accept. Even when the stakes are high, it remains epistemically impossible for me that my bike is red and blue all over.

When p is epistemically possible for an agent, we can say that there is an epistemically possible *scenario* for the agent where p is true.[3] Intuitively, we can think of a scenario as a maximally specific way things might be. Corresponding to the epistemic possibility that there is extraterrestrial life, there will be various scenarios in which green men dance on Mars and in which crabs crawl on Jupiter. And corresponding to the epistemic possibility that Fermat's Last Theorem is false, there will be various scenarios in which there are integers a, b, c and n greater than 2 that satisfy the equation $a^n + b^n = c^n$.

We can think of the space of all such scenarios as *epistemic space*. Relative to each agent at a particular time, a class of scenarios in epistemic space are singled out as epistemically possible for the agent. For instance, I know that the sun shines in Canberra, but you do not. Whereas the sun shines in Canberra in all scenarios that remain epistemically possible for me, the rain is pouring down in Canberra in some of the scenarios that remain epistemically possible for you. I then tell you that the sun shines in Canberra. You come to know the proposition and rule out all scenarios where it rains in Canberra as epistemically impossible. More generally, for a given proposition p, we can say that there are scenarios in epistemic space where p is true, and scenarios in epistemic space where p is false.[4] When an agent comes to know p, all scenarios

[3] The terms 'scenario' and 'epistemic space' below are borrowed from Chalmers (forthcoming). For similar terminology, see also Hintikka (2003).

[4] A word of warning: The use of propositions in the introductory material might offend people with firm opinions on the nature of propositions. In chapter 2, I will set aside propositions altogether and associate epistemic possibility with sentences. But it is easier to motivate the general picture in terms of propositions as a "placeholder" notion.

in epistemic space where p is false are ruled out as epistemically impossible for the agent. But of course these scenarios may well remain epistemically possible for agents that do not know p.

A picture similar to the one I have sketched here can be found in many branches of epistemology. It captures the relationship between *gaining* information and *excluding* possibilities that constitutes the backbone of standard possible world models of belief and knowledge.[5] I have illustrated how this basic framework works for knowledge and epistemic possibility, but it applies straightforwardly to belief and doxastic possibility as well. When p might be the case for all an agent believes, we can say that p is *doxastically possible* for the agent. When an agent comes to believe that p, all scenarios in epistemic or doxastic space where p is false are ruled out as doxastically impossible for the agent. Arguably, since belief is a more fundamental notion than knowledge, doxastic possibility is a more fundamental notion than epistemic possibility. Yet, since epistemic possibility is the more familiar player in the literature, and since the main insights of this project apply equally well to doxastic and epistemic possibility, I will continue to phrase things in terms of epistemic possibility.

Though the basic ideas behind the intuitive picture are familiar, the details remain difficult. Particularly acute is the question of what *scenarios* are. To make the basic task of this project clear, I will first analyze this central notion in more detail.

1.2 Knowledge, Epistemic Possibility and Scenarios

Epistemic possibility is related to what different agents know and do not know. In particular, the current notion of epistemic possibility applies to agents

[5] Refer, amongst many others, to Barwise (1997), Dretske (1981), Hintikka (1962), Jackson (1998), Lewis (1986), and Stalnaker (1984).

that may fail to know that water is H_2O and that Fermat's Last Theorem is true. To make the relationship between knowledge and epistemic possibility more precise, let us adopt the following analyses:[6]

(**EP***) A proposition p is epistemically possible for an agent a when a cannot easily come to know $\neg p$ from what a already knows.

(**EN***) A proposition p is epistemically necessary for an agent a when a can easily come to know p from what a already knows.

Though I will be more explicit about the details in chapter 2, I will assume that (EP*) and (EN*) serve as plausible analyses of the (philosophically) ordinary notions of epistemic possibility and necessity. Intuitively, even if a chemically ignorant agent knows that water is water, there is no easy way for her to come to know that water is H_2O from that piece of knowledge. And intuitively, even if a competent mathematician knows the basic principles and axioms of number theory, there is no easy way for her to come to know that Fermat's Last Theorem is true from these pieces of knowledge. For now, we can think of the relevant kind of reasoning that figures in (EP*) and (EN*) as armchair reasoning. Roughly, if an agent can easily come to know p by reasoning from what she already knows, then *that* chain of reasoning may involve empirical information that the agent already has, but not any further empirical information.

To make the relationship between epistemic possibility and scenarios more precise, let us first follow Chalmers and distinguish between *strict* and *deep* epistemic possibility and necessity:

> [T]he notion of *strict epistemic possibility*—ways things might be, for all we know—is undergirded by a notion of *deep epistemic possibility*—ways things might be, prior to what anyone knows. Unlike strict epistemic possibility, deep

[6] See DeRose (1991), Huemer (2007), Stanley (2005), and Teller (1972) for similar analyses. The star $\star$ in (EP*) and (EN*) indicate that these analyses are in need of slight adjustments. The adjustments will be made in chapter 2.

> epistemic possibility does not depend on a particular state of knowledge, and is not obviously relative to a subject. Whereas it is strictly epistemically possible (for a subject) that p when there is some epistemically possible scenario (for that subject) in which p, it is deeply epistemically possible that p when there is some deeply epistemically possible scenario in which p.[7]

As I will think of the distinction, strict epistemic possibility and necessity correspond to the notions of epistemic possibility and necessity that figure in (EP*) and (EN*). Though it is strictly epistemically impossible for me that it is raining in Canberra, it nevertheless remains deeply epistemically possible that it is raining in Canberra. When a given agent comes to know p, this piece of knowledge divides the class of deeply epistemically possible scenarios into those scenarios where p is true and into those scenarios where p is false. If an agent knew nothing, all deeply epistemically possible scenarios would also be strictly epistemically possible for this agent.

We can also put this by saying that whereas the class of deeply epistemically possible scenarios constitute epistemic space W, the class of scenarios in W that remain strictly epistemically possible for a given agent a (at a given time) constitute strict epistemic space W_a. The relationship between strict and deep epistemic possibility can then be captured by saying that for any agent a, $W_a \subseteq W$. So if w is in W_a, then w is in W, though the converse need not be the case. In this sense, deep epistemic possibility is a necessary, yet not sufficient condition for strict epistemic possibility. Intuitively, the notion of deep epistemic possibility delineates the borders of epistemic space, and within it, strict epistemic possibility for a delineates the borders of what a knows. Henceforth, all unqualified talk of epistemic possibility refers to strict epistemic possibility.

[7] Chalmers (forthcoming): p. 4.

Given this, we can then state the basic relationship between epistemic possibility, necessity and scenarios as follows:

(**Epi-Pos***) A proposition p is epistemically possible for an agent a just in case there is a scenario w in W such that w is epistemically possible for a and such that p is true at w.

(**Epi-Nec***) A proposition p is epistemically necessary for an agent a just in case for each scenario w in W such that w is epistemically possible for a, p is true at w.

These analyses are motivated by Hintikka (1962), and together with (EP*) and (EN*) they capture the central idea behind world involving analyses of epistemic notions such as belief and knowledge. If an agent knows p, then p is true at all scenarios w in W that are epistemically possible for this agent.

We now have the basic ingredients to shed more light on the question of what scenarios are. I am interested in agents that may know that water is water, but not that water is H_2O, and in agents that may know that $2+2=4$, but not that Fermat's Last Theorem is true. We want scenarios in W such that (Epi-Pos*) and (Epi-Nec*) are plausible principles for these kinds of agents. This means admitting scenarios in W where water is not H_2O and where Fermat's Last Theorem is false. If we need these kinds of scenarios in W, it is easy to see what scenarios *cannot* look like.

First, assume we think of scenarios as metaphysically possible worlds, and assume that we identify the class of scenarios W in (Epi-Nec*) with the class of metaphysically possible worlds. Then (Epi-Nec*) reads as:

(**Epi-Nec**$^\star_{\mathbf{M}}$) A proposition p is epistemically necessary for an agent a just in case for each metaphysically possible world w in W such that w is epistemically possible for a, p is true at w.

Let p be the proposition that water is H_2O. On a standard conception of metaphysically possible worlds, p is true at all metaphysically possible worlds. Accordingly, no matter which class of metaphysically possible worlds remain epistemically possible for any given agent, p will be true at each such world. By (Epi-Nec$^{\star}_{\mathrm{M}}$), p is then epistemically necessary for any agent. But since there are agents that do not know that water is H_2O, p cannot be epistemically necessary for all agents. So (Epi-Nec$^{\star}_{\mathrm{M}}$) is false, and scenarios cannot be metaphysically possible worlds.[8]

Second, assume we aim to think of scenarios as epistemically possible worlds. Though there are several ways to make the loose and intuitive notion of an epistemically possible world precise, I will use elements from Chalmers (forthcoming) to motivate broadly ersatz constructions of these entities as, roughly, maximal a priori consistent sets of sentences or propositions.[9]

For Chalmers,

[8] Strictly, this conclusion is too hasty. Based on his causal-pragmatic account of intentionality, Stalnaker (1984) argues that agents *are* in fact omniscient with respect to all (metaphysically) necessary truths. To accommodate our intuitions to the contrary, Stalnaker advances his metalinguistic or diagonalization strategy. Roughly, if an agent *seemingly* fails to know that water is H_2O, what she really fails to know is the contingent proposition that the *string* 'Water is H_2O' expresses the necessary proposition. Since there are plenty of metaphysically possible worlds in which the string 'Water is H_2O' means that cockatoos eat magpies, the rough idea behind Stalnaker's view is to locate apparent ignorance of a necessary proposition in ignorance of an associated metalinguistic contingent proposition.

There is no shortage of critiques of Stalnaker's combined fragmentation and metalinguistic strategy; see for instance Field (2001): chapter 3, Lycan (1990), and Robbins (2004). And although an adequate discussion of views like Stalnaker's is beyond the scope of this project, I will briefly add the following worry. On Stalnaker's view, there is no in principle difference between a possible world w that falsifies the proposition p that ['Two plus two equals four' expresses the necessary proposition], and a world w_1 that falsifies the proposition p_1 that ['There are no integers a; b; c; $n > 2$ such that $a^n + b^n = c^n$' expresses the necessary proposition]. But if we look for an explanation of the striking fact that not even the greatest mathematicians until recently could rule out w_1, while they could easily rule out w—alongside most other moderately ideal agents—we should expect an in principle difference between worlds like w and worlds like w_1. Since such an in principle difference is missing in Stalnaker's account, it ends up bearing resemblance to the "anything-goes" construction of Extreme Epistemic Space that I will investigate in chapter 2. And for reasons that will become clear, such a construction is unsuited for many purposes—and notably for modeling moderately ideal agents.

[9] See Chalmers (forthcoming) for all the details. For other conceptions of epistemically possible worlds, see Soames (2005).

> [. . .] s is deeply epistemically necessary when s is *a priori*: that is, when s expresses actual or potential a priori knowledge. More precisely, s is a priori when it expresses a thought that can be justified independently of experience, yielding a priori knowledge.[10]

A proposition p is then deeply epistemically possible when $\neg p$ is not deeply epistemically necessary. For Chalmers,

> [t]his idealized notion of apriority abstracts away from contingent cognitive limitations. If there is any possible mental life that starts from a thought and leads to an a priori justified acceptance of that thought, the thought is a priori. [. . .] So if a hypothesis can be known to be false only by a great amount of a priori reasoning, it is nevertheless deeply epistemically impossible. For example, 'There are integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$' is deeply epistemically impossible.
>
> [. . .]
>
> When apriority is understood as above, it is clear that typical tokens of sentences such as 'Hesperus is Phosphorus' are not a priori. The thoughts expressed by these tokens are such that there is no possible mental life that starts from that thought and leads to an a priori justified acceptance of that thought.[11]

Given this notion of deep epistemic possibility, Chalmers goes on to construct scenarios as equivalence classes of epistemically complete sentences in an ideal language $\mathcal{L}^+$.[12] A sentence S in $\mathcal{L}^+$ is *epistemically complete* when S is deeply epistemically possible, and there is no other sentence T in $\mathcal{L}^+$ such

[10] Chalmers (forthcoming): p. 7. To avoid making any presuppositions about the nature of propositions, Chalmers only works with sentences and thoughts. For now, however, I will simply assume that his notion of deep epistemic possibility applies to the "placeholder" notion of propositions at work here. In chapter 2, I return to these issues.

[11] Chalmers (forthcoming): pp. 8-9.

[12] In chapter 2, I return and discuss "scenario-making" languages in more details, but for now I refer the reader to Chalmers (forthcoming): pp. 17-22. Chalmers also argues that *centered possible worlds*, when properly understood, can do the relevant work that he wants scenarios to do. Here I will only focus on the linguistic construction he offers.

that $(S \wedge T)$ and $(S \wedge \neg T)$ are both deeply epistemically possible. Two epistemically complete sentences S and T are *equivalent* just in case S implies T and T implies S, where S *implies* T when $(S \wedge \neg T)$ is deeply epistemically impossible; that is, when $(S \rightarrow T)$ is deeply epistemically necessary. A scenario w then corresponds to an epistemically complete sentence in the sentence's equivalence class.

We can think of Chalmers' construction of scenarios as formalizing the intuitive notion of an epistemically possible world. And we can think of the basic material in Chalmers' construction as explicating the core idea behind linguistic constructions of epistemically possible worlds as maximal, a priori consistent sets of sentences in a scenario-making language. A priori consistent in the sense that no conjunction of sentences in the set is deeply epistemically impossible. And maximal in the sense that the conjunction of sentences in the set with any sentence outside the set is deeply epistemically impossible. More generally, we can use these rough details to characterize the structural features of what I will call *Ideal Epistemic Space.*

Let us say that p is *ideally deeply epistemically necessary* when p is a priori, where the notion of apriority is understood in the idealized sense above. With a view to Chalmers' construction of scenarios, let us then assume that we have a construction of a space of scenarios that is grounded in this notion of ideal deep epistemic necessity. Then for any ideally deeply epistemically necessary p, p is true at each scenario w in this space. Call these scenarios *ideal scenarios*, and call the space of ideal scenarios *Ideal Epistemic Space.*

Given this, suppose we think of epistemically possible worlds as ideal scenarios, and suppose we identify the class of scenarios W in (Epi-Nec*) with the class of ideal scenarios. Then (Epi-Nec*) reads as:

(**Epi-Nec$^\star_{\mathbf{I}}$**) A proposition p is epistemically necessary for an agent a

> just in case for each ideal scenario w in W such that w is epistemically possible for a, p is true at w.

Let p be the proposition that water is H_2O. Since p is not ideally deeply epistemically necessary, there are ideal scenarios in W where p is false. So p need not be true at each ideal scenario that remains epistemically possible for an agent. By (Epi-Nec$^\star_\mathrm{I}$), then p need not be epistemically necessary for any agent, and hence agents may fail to know that water is H_2O. The problem with (Epi-Nec$^\star_\mathrm{M}$) has been solved. So in contrast to constructions of W based on metaphysically possible worlds, constructions of W based on ideal scenarios or epistemically possible worlds allow us to model agents that do not know a posteriori necessities such as 'Water is H_2O' and 'Hesperus is Phosphorus'.

But now let p be the proposition that Fermat's Last Theorem is true. Since p is ideally deeply epistemically necessary, p is true at all ideal scenarios in W. Accordingly, no matter which class of ideal scenarios remain epistemically possible for any given agent, p will be true at each such scenario. By (Epi-Nec$^\star_\mathrm{I}$), then p is epistemically necessary for any agent. But since there are agents that do not know that Fermat's Last Theorem is true, p cannot be epistemically necessary for all agents. So (Epi-Nec$^\star_\mathrm{I}$) is false, and scenarios cannot be ideal scenarios.

As a result of defining deep epistemic possibility in terms of the idealized notion of apriority, as Chalmers notes, Ideal Epistemic Space "is best suited for modeling the knowledge and belief of idealized reasoners that may be empirically ignorant, but that can engage in arbitrary a priori reasoning."[13] Since I am interested in non-ideal agents that may not only be empirically ignorant, but also only have limited cognitive capacities available for reasoning, complex a priori falsehoods can remain epistemically possible for such agents. Accordingly, to make (Epi-Pos*) and (Epi-Nec*) plausible principles for the broad

[13] Chalmers (forthcoming): p. 8.

class of agents that are not ideal, we need go to beyond Ideal Epistemic Space.

Though I have focused on complex mathematical truths in the discussion of ideal scenarios, similar problems obviously arise for complex logical falsehoods. Not only are there many complex mathematical truths but also many complex logical truths that non-ideal agents can fail to know. By the line of reasoning above, this immediately implies that scenarios cannot be identified with logically possible worlds, maximally consistent sets of sentences or propositions, or their close relatives.[14] Rather, to make (Epi-Pos*) and (Epi-Nec*) plausible principles for agents that are not mathematically nor logically omniscient, we need scenarios in W at which mathematical and logical truths can be false. Using Hintikka's intuitive gloss "[t]his means admitting 'impossible possible worlds', that is, worlds which look possible and hence must be admissible as epistemic alternatives but which none the less are not logically possible."[15] Call scenarios at which a priori necessary truths can be false *non-ideal scenarios*, and call epistemic spaces that contain non-ideal scenarios *non-ideal epistemic spaces*.[16]

In light of this, the basic project of the dissertation is to construct and understand non-ideal scenarios so that (Epi-Pos*) and (Epi-Nec*) are plausible principles for the broad class of agents that are not ideal.

1.3 Non-Ideal Epistemic Space

In investigating subsequent constructions of non-ideal epistemic spaces, I will follow Chalmers' general approach along two lines.

First, I will base various constructions of non-ideal epistemic space on var-

[14] Unless, of course, we define consistency with respect to a logic that has no theorems, but I set aside such cases.

[15] Hintikka (1975): p. 477.

[16] Since I will not discuss issues that might arise with respect to *contingent* a priori truths, I will usually just say 'a priori truths' intending this to mean 'a priori necessary truths' such as those of mathematics and logic.

ious *non-ideal* notions of deep epistemic possibility. Since we need to ensure that not all a priori truths are epistemically necessary for non-ideal agents, we need a non-ideal notion of deep epistemic possibility, according to which not all a priori truths are deeply epistemically necessary. In chapter 2, I investigate a notion according to which any sentence, and in particular any a priori false sentence can be deeply epistemically possible. In chapter 3, I then investigate a notion according to which some, but not all a priori false sentences can be deeply epistemically possible.

Second, I will *construct* scenarios. Since non-ideal scenarios are akin to impossible worlds, a linguistic or ersatz construction of these entities is the most natural approach. We already have a rather good grip on what it means to identify possible worlds with maximal, consistent sets of propositions or interpreted sentences in some world-making language. In this camp, we find Adams' complete consistent sets of propositions, Carnap's state descriptions, Chalmers' equivalence classes of epistemically complete sentences, and Hintikka's and Jeffrey's complete consistent novels.[17] Insofar as we can give an explicit construction of scenarios as maximal sets of sentences, some of which may fail to be consistent in some relevant sense of consistency, then non-ideal scenarios seem to deserve the label *world-like entities*. As Nolan notices:

> For most abstractionists [or ersatzers about possible worlds], in fact, it would seem that accepting impossible worlds, and even impossibilia, would be only accepting ontology of a sort which they are already committed to. In some cases, they would not even need to accept anything new: someone who took possible worlds to be sets of propositions, or sets of sentences-like representations, is probably already committed to sets of sentences which are not [...] consistent[.] These other sets may well represent perfectly adequately ways the world could not turn out.[18]

[17] See Adams (1974), Carnap (1947), Chalmers (forthcoming), Hintikka (1962), Hintikka (1969), and Jeffrey (1983).

[18] Nolan (1997): p. 542. Of course, if ones aims to understand impossible worlds from

For purposes of this project, non-ideal scenarios will earn their keep by the explanatory roles that they can play. And as we shall see now, they can potentially play many useful roles.

1.4 Hyperintensional Phenomena

If we can set up a non-ideal epistemic space successfully, we can use scenarios in this space to avoid many of the familiar hyperintensional problems that emerge in standard possible world frameworks. In illustrating these hyperintensional problems, I will take the standard possible world framework to be committed to the claim that all logical and mathematical truths are true at each possible world. With the possible exception of mathematical truths, this is a fair characterization of what is common among different conceptions of possible worlds.[19]

In philosophy of language, it is standard to use *intensions*, which are functions from possible worlds to extensions, to help us understand the notion of meaning. But it is also well-known that such *possible world intensions* do not seem to capture our linguistic intuitions in many situations. The standard examples involve Fregean puzzles and attitude ascriptions. To illustrate, consider any two mathematically equivalent sentences A and B. Since A and B are true at all possible worlds, A and B are co-intensional. So A and B have the same meaning or semantic content. If language is compositional, we should be able to substitute A and B in any sentence that contains either as a constituent without changing the truth-value of the resulting sentence. But there are many

the perspective of a Lewisian modal realism, things will look much more complicated; see Yagisawa (1988). For other views on the metaphysical nature of impossible worlds, see Vander Laan (1997) and Zalta (1997).

[19] For instance, if one has reason to believe that all mathematical truths are contingent, one might also have reason to settle for a construction of possible worlds that allows that mathematical truths can be false at these worlds. Though I could easily rephrase all the examples below in terms of logical truths, mathematics provides for good and intuitive cases. So for now I will assume that mathematical truths obtain in all possible worlds.

mathematically equivalent A and B for which it can be intuitively true to say things like ‘Julie believes A’ but false to say things like ‘Julie believes B’. If A is the sentence ‘$2 + 2 = 4$’ and B is the sentence ‘There are no integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$’, we have a case. If so, there seems to be more to meaning that mere intensional content.

Similar motivation comes from analyses of counterfactual conditionals with impossible antecedents. Borrowing a case from Nolan, the counterpossible ‘If Hobbes had squared the circle, sick children in the mountains of South America at the time would not have cared’ seems true, while the counterpossible ‘If Hobbes had squared the circle, then everything would have been the case’ seems false.[20] But according to the standard Lewis-Stalnaker semantics for counterfactuals, counterpossibles are always true. Since the intension of the antecedent in both counterpossibles is false at all possible worlds, then vacuously all the closest worlds in which the antecedent is true are worlds in which the consequent is true. But intuitively the two counterpossibles differ in truth-value. If so, there seems to be more to meaning that mere intensional content.

In philosophy of mind, many people have used possible world intensions to represent the content of propositional attitudes like belief and knowledge. With a view to the discussion above, we can represent the content of an agent’s belief that p by the class of possible worlds where p is true. Yet, since the class of possible worlds where $2 + 2 = 4$ just is the class of possible worlds where Fermat’s Last Theorem is true, we have to represent an agent that believes $2 + 2 = 4$ as thereby also believing that Fermat’s Last Theorem is true. But intuitively, an agent can believe the former without the latter. If so, there seems to be more to propositional content than mere intensional content.

[20] Cf. Nolan (1997): p. 544. The same considerations can be brought to bear on world involving analyses of indicative conditionals.

In epistemology, a closely related problem goes under the label *the problem of logical omniscience*. The problem arises in many of our best formal theories of belief, knowledge and information. In a standard Bayesian framework, credences are normally distributed over a set of possible worlds. If so, then agents are modeled as assigning credence 1 to all logical and mathematical truths. In standard doxastic and epistemic logics, belief and knowledge are represented by modal operators that receive a standard Kripke semantics in terms of quantification over possible worlds. If so, then agents are modeled as knowing and believing all logical and mathematical truths. But if we want to use Bayesian epistemology and epistemic logic to illuminate facts about ordinary reasoners, we cannot distribute credences over nor analyze knowledge merely in terms of possible worlds.[21] A variation of the problem of logical omniscience is known as "the scandal of deduction".[22] In standard possible world models, the informational content of a proposition p can be represented by the (finite) set of possible worlds where p is false. However, since logical and mathematical truths are true at all possible worlds, they strictly have no informational content. Yet intuitively, we often can and do gain new information from deductive reasoning, and it is a "scandal" if our formal theories tell us otherwise.[23] So if we want to use our epistemological theories to illuminate the ordinary notions of belief, knowledge and information, it seems that we need to appeal to more than classes of possible worlds.

These considerations suggest that philosophy of language, philosophy of mind and epistemology need a notion of content that is more fine-grained than

[21] As Hintikka says:

> Since the assumption of [...] logical omniscience is obviously mistaken, this commitment seems to constitute a grave objection to the whole possible-worlds treatment of propositional attitudes. (Hintikka (1989): p. 63.)

[22] Cf. Hintikka (1973): p. 222.

[23] This is known as the Bar-Hillel-Carnap paradox; see Bar-Hillel and Carnap (1953). For recent discussions, see D'Agostino and Floridi (2009) and Sequoiah-Grayson (2008).

traditional intensional content. Hyperintensional content is supposed to fill this role. More specifically, if we have impossible worlds or non-ideal scenarios in modal space, we can aim to define *hyperintensions* over these more fine-grained possibilities. In contrast to the standard intensions above, hyperintensions will be functions from non-ideal scenarios to a truth-value. Assuming that we have a well-defined class of non-ideal scenarios and hyperintensions in our toolbox, we can then run through a few examples to show the potential work that they might do.

First, we can attempt to understand semantic content in terms of hyperintensions. For instance, though the standard intensions of 'Julie believes A' and 'Julie believes B' coincide in truth-value for any two mathematically equivalent sentences A and B, the values of the hyperintensions of these two sentences need not. Since the mathematical truth B, for instance, can be false at non-ideal scenarios, the values of the hyperintensions of 'Julie believes A' and 'Julie believes B' can come apart. Similarly, though the intuitively false counterpossible 'If Hobbes had squared the circle, then everything would have been the case' is vacuously true when modal space is exhausted by possible worlds, it may well be false when the hyperintension of the antecedent can be true at non-ideal scenarios. In this sense, hyperintensions can potentially play the role of a fine-grained notion of semantic content in philosophy of language.

Second, we can attempt to use hyperintensions to represent the contents of thoughts and beliefs. For instance, the hyperintensions of the thoughts that $2 + 2 = 4$ and that Fermat's Last Theorem is true need not coincide in truth-value when evaluated at non-ideal scenarios. So we can represent an agent as believing that $2 + 2 = 4$ without thereby also representing the agent as believing that Fermat's Last Theorem is true. In this sense, hyperintensions can potentially play the role of a fine-grained notion of mental content in philosophy of mind.

Third, we can attempt to use the class of non-ideal scenarios as the underlying space of possibilities that figures in various world models in epistemology. For instance, we might aim to analyze the knowledge operator K in epistemic logic in terms of quantification over non-ideal scenarios. Roughly, even if $K(A)$ is true at some scenario w, then some B that is logically equivalent to A might be false at some non-ideal scenario w' that is epistemically accessible from w. If so, then $K(B)$ need not be true at w, in which case knowledge need not be closed under logical equivalence.[24] Or we can attempt to distribute credences over non-ideal scenarios. Roughly, we could then allow that $Cr(A) < 1$, where Cr is a credence function and A is a logical truth. If so, then agents need not be modeled as assigning credence 1 to all logical truths. In this sense, non-ideal scenarios can potentially play a role in isolating a fine-grained space of possibilities that we can use to model agents that are not logically omniscient.

Given these examples, I trust that the explanatory roles that non-ideal scenarios can potentially play are many and useful. It should be noted that the impossible world approach to hyperintensionality is not the only one. We can isolate two broad alternatives. First, we have approaches that appeal to a notion of structured content, which either determines or supplements the intensional notion of content. For instance, we can appeal to a notion of Russellian content that allows us to say that sentences and thoughts about mathematical and logical truths can have different content. Such Russellian content can then either determine intensional content, or it can be combined with intensional content to form a complex content of some sort.[25] Second, we

[24] For detailed discussion of semantics for epistemic logic that involve impossible worlds, see Fagin et al. (1995): pp. 357-362 and Wansing (1990).

[25] See, for instance, Soames (1987). On a Perry-style view on beliefs, the relation between belief and content or (Russellian) proposition is roughly mediated by something akin to modes of presentations or guises that are tied to psychological features of the believer; see for instance Crimmins and Perry (1989). If such guises are individuated in a sufficiently fine-grained manner, it seems in principle possible to use them to handle most, if not all hyperintensional problems. If so, we might include a Perry-style view on the metaphysics of beliefs as a third distinct approach to hyperintensionality.

have approaches that aim to reconcile the standard possible world framework with hyperintensional phenomena. Famously, Stalnaker has argued that many hyperintensional phenomena, when properly understood, can be captured using just intensional content. On Stalnaker's view, propositional content is exhausted by possible world content, and apparent hyperintensional problem cases are explained away using a combined fragmentation and metalinguistic strategy.[26]

I will not attempt to evaluate and compare these alternatives to the impossible world route that I will investigate in this project. But *if* we bracket a Stalnakerian reconciliation approach, as I will, let me briefly mention two general motivations for going the impossible world route. First, *if* we can construct a non-ideal epistemic space that enjoys nice formal and intuitive properties, we might use scenarios in this space to motivate a notion of propositional content, which although unstructured is as fine-grained as structured propositional content. Derivatively, we can use this notion to make progress on the hyperintensional problems above. Second, *if* we can construct a non-ideal epistemic space that enjoys nice formal and intuitive properties, we might have a general purpose space of scenarios that can potentially feed in to various world involving models in epistemology. Closest to home, we might be able to use such a space to capture the intuitive picture of epistemic possibility outlined above. But more generally, *if* non-ideal epistemic space has an interesting formal structure, as the traditional Boolean structure that underlies standard possible world frameworks, we might attempt to investigate non-classical probabilistic and logical models that are based on this non-ideal epistemic space.

[26] See in particular Stalnaker (1984); see also footnote 8, page 9.

1.5 Non-Trivial Epistemic Spaces

As we shall see in chapter 2, it is rather simple to construct a non-ideal epistemic space in which we can avoid all hyperintensional problems. Scenarios in this space, which I will call *Extreme Epistemic Space*, are akin to Priest's *open worlds* in which arbitrary logical contradictions and inconsistencies can be true.[27] For reasons that I will only briefly indicate at this point, I will mainly be interested in constructions of *non-trivial epistemic spaces.* Very roughly, a non-trivial epistemic space is a space of scenarios where not "anything goes". Less roughly, scenarios in a non-trivial epistemic space will obey certain substantive constraints. The constraints that I want scenarios in non-trivial epistemic space to obey will in general be motivated by two desiderata.

Ideally, the first desideratum goes, we should be able to use scenarios in non-ideal epistemic space to give a world involving analysis of a non-trivial notion of hyperintensional content. Call this the *content desideratum.* Ideally, the second desideratum goes, we should be able to use scenarios in non-ideal epistemic space to give a world involving analysis of a non-trivial notion of epistemic possibility that captures which propositions *should* and *should not* remain epistemically possible for non-ideal agents. Call this the *rationality desideratum.* Here I will motivate each desideratum and indicate how they provide prima facie motivation for investigating constructions of non-trivial epistemic spaces. In chapters 2 and 3 I elaborate further on the details.

I will think of the notion of *non-triviality* that figures in both desiderata as tied to a notion of non-ideal reasoning of a particular kind. To illustrate, I borrow an example from Cherniak.[28] Intuitively, if I know the proposition p that there are 2 apples in that basket and 3 apples in this basket, I can easily come to know the proposition q that there are 5 apples. On the other hand,

[27] See Priest (2005).

[28] Cf. Cherniak (1986): p. 29. See also Barwise (1997) for similar examples.

if I know the proposition p_1 that there are 298 apples in each of 783 baskets, I cannot easily come to know the proposition q_1 that there are 233,334 apples. In the first case, if I know p, then there is a chain of arithmetical reasoning that I could easily, obviously, or instantly perform to infer q. Intuitively, if I were to reflect on p, I could immediately come to accept q. Not so for the complex chain of arithmetical reasoning that is involved in the second case. If I know p_1, there is no chain of arithmetical reasoning that I could easily, obviously, or instantly perform to infer q_1. Intuitively, if I were to reflect merely on p_1, I could not immediately come to accept q_1.

If this intuitive picture is correct, and I will assume that it is, then a notion of hyperintensional content or epistemic possibility can be said to be non-trivial if it reflects these basic non-trivial inferential relations among various thoughts and beliefs. For agents that have different cognitive capacities available for easy, obvious, or instant reasoning, different inferential relations among thoughts and beliefs will count as trivial. For an ideal agent that can engage in arbitrary a priori reasoning, the inference from the thought that p_1 to the thought that q_1 will be trivial. And maybe there are certain agents, the extremely non-ideal ones, for whom the inference from the thought that p to the thought that q will be non-trivial. For the broad class of moderately ideal agents, at least certain basic inferential relations among thoughts will count as trivial—for instance the inference that leads from the thought that there are 2 apples in that basket and 3 apples in this basket to the thought that there are 5 apples.

1.5.1 Content Desideratum

For the content desideratum, we want to use scenarios in non-ideal epistemic space to give a world involving account of a non-trivial notion of hyperintensional content. In a world involving analysis of thought and belief content, we aim to represent content in terms of classes of possibilities. To represent the content of non-ideal epistemic states, we can aim to use non-ideal scenarios. In particular, we can aim to represent the contents of non-ideal epistemic states by *non-ideal epistemic intensions*, which will be functions from ideal and non-ideal scenarios to truth-values.[29] The non-ideal epistemic intension of a belief with a certain content p will map a non-ideal scenario to the truth-value of p in that scenario. With a view to the intuitive picture from above, the content of a belief will correspond to the way it divides non-ideal epistemic space.

Suppose then that we have a well-defined space of non-ideal scenarios, and suppose we want to use non-ideal epistemic intensions to represent the contents of the epistemic states of moderately ideal agents. Since the contents of such epistemic states can stand in non-trivial inferential relations to each other, we ideally want to reflect these relations in the non-ideal epistemic intensions that we use to model these contents. To use the example from above, this would mean that the non-ideal epistemic intension of the thought that q should be true at a scenario w just in case the non-ideal epistemic intension of the thought that p is true at w. But if so, then this suggests that we need a construction of non-ideal scenarios that obey certain constraints and where not "anything goes". If everything went, there might well be non-ideal scenarios where p is true and q is false, in which case we cannot use *these* scenarios to define the kinds of non-trivial, yet non-ideal epistemic intensions that are appropriate for modeling the contents of the epistemic states of moderately ideal agents.

In this sense, *if* we can give a successful construction of non-trivial epistemic space, we might use it to develop an interesting world involving analysis of a non-trivial notion of hyperintensional content. Derivatively, we could use such an analysis to do substantial work in philosophy of language, mind and epistemology.

[29] The term 'non-ideal epistemic intension' is taken from Chalmers (forthcoming).

In particular, following Chalmers, there is hope that non-ideal epistemic intensions may behave as a sort of Fregean content.[30] The Fregean content of a thought or an expression is its *sense*, where sense is an aspect of content that is tied to cognitive significance. Roughly, two thoughts or sentences A and B have the same sense just in case $A \leftrightarrow B$ is cognitively insignificant. We can then first attempt to analyze *sense* in terms of non-ideal epistemic intensions. Second, we can attempt to say that a thought is *cognitively insignificant* when a moderately ideal agent can come to known or accept the thought by easy, obvious or instant a priori reasoning—or, by invoking the analysis of epistemic possibility above, whenever the thought is epistemically necessary for any moderately ideal agent, independently of what empirical information she might happen to have. This naturally motivates an analysis, according to which A and B have the same non-ideal epistemic intension just in case $A \leftrightarrow B$ is epistemically necessary for any moderately ideal agent. Insofar as the non-ideal epistemic intensions of logical and mathematical truths can come apart, we could then vindicate the Fregean idea that logically and mathematically equivalent thoughts and expressions can have different senses.

This provides motivation for the content desideratum, and derivatively prima facie motivation for investigating non-trivial epistemic spaces.

1.5.2 Rationality Desideratum

For the rationality desideratum, we want to use scenarios in non-ideal epistemic space to give a world involving analysis of a non-trivial notion of epistemic possibility that captures which propositions *should* and *should not*

[30] See Chalmers' work on the *Golden Triangle*, which promises to restore the Kantian link between reason and modality, the Fregean link between reason and meaning or content, and the Carnapian link between meaning and modality; cf. Chalmers (2004). See Chalmers (2002a) for details on the relation between epistemic intensions and Fregean content, and see also Chalmers (2002b) for some of the many puzzles that epistemic intensions could potentially resolve.

remain epistemically possible for agents that are not extremely non-ideal. The basic observation is this: Just because we give up the requirement that all agents are ideally rational, we still "ascribe knowledge, belief, the ability to form judgements and so on only to *agents*, i.e., systems which we take to be (or which can be interpreted as) rational to some degree."[31] At least for a project in epistemology, this claim seems platitudinous.

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. One can, therefore, hold that the philosophically interesting notion of epistemic space is a rational space, incorporating the normative element of our epistemic concepts, without thereby holding it to be an ideal epistemic space in Chalmers' [Ideal Epistemic Space] sense.
>
> Our concept of epistemic possibility (and hence of an epistemic scenario) is a normative concept which, at the same time, should allow that valid inferences can be informative.[32]

The guiding idea behind this conception of rationality is simple: If we bracket aspects of non-ideal reasoning that pertain to mistakes, forgetfulness and lack of attention, non-ideal reasoning is not fundamentally different from ideal reasoning. It is just limited. If we say that these limitations are determined by the cognitive resources that agents have available for reasoning, we can naturally invoke the following minimal requirement of rationality: If an agent can

[31] Jago (2009a): p. 332.

[32] Jago (2009a): p. 333. Recently, resource-bounded reasoning and rationality have been investigated intensively in artificial intelligence and formal epistemology; see Jago (2006) for a broad overview and further references.

come to believe *p* by easy, obvious or instant reasoning from what she already believes, then the agent rationally *should* believe *p*. A minimally rational or moderately ideal agent is then an agent that never violates this requirement.[33]

The (EP*) and (EN*) analyses, which arguably capture the core of standard conceptions of epistemic possibility and necessity, nicely reflect the idea behind resource-bounded rational reasoning. With a view to (EN*), if an agent *can* easily come to know *p* from what she already knows, then *p* *is* epistemically necessary for this agent.[34] Whether or not ordinary reasoners *actually* engage in the relevant kind of easy reasoning that leads to *p*, or whether their thoughts are distracted, confused or not focused on *p*, the dispositional capacity to engage in this reasoning remains. In this sense, (EP*) and (EN*) capture which propositions *should* remain epistemically possible and necessary for ordinary, reflective reasoners. That is, they capture which propositions are epistemically possible and necessary for minimally rational agents that always engage in the relevant kind of reasoning that ordinary, reflective reasoners can easily, and often do engage in. This gives the (EP*) and (EN*) analyses a normative twist that reflects the normative component of our concept of epistemic possibility.

Notice, however, that this normative component need not prevent us from saying that there are extremely non-ideal agents for whom just about *any* proposition is epistemically possible. We will just have to stipulate that such agents have no, or only few cognitive capacities available for easy reasoning. In

[33] Along the lines of Cherniak's "no rationality, no agent" dictum, one might hold that a notion of minimal rationality plays a *constitutive* role for what it means to be a person with beliefs and desires; cf. Cherniak (1986). Arguably, the current minimal notion of rationality can play such a constitutive role. Intuitively, since the notion can be motivated by reference to those inferences that cognitive systems like ours are always disposed to make, it is hard to imagine how our thoughts and beliefs could violate this minimal requirement wildly. For recent discussions of the constitutive role of rules of rationality, see Boghossian (2003), Glüer and Wikforss (2009), and Wedgwood (1999).

[34] Huemer (2007) complains that analyses along the lines of (EP*) and (EN*) are insufficiently informative to be satisfying. In chapter 3, I will provide a precise test case interpretation of 'easy reasoning', which at least provides us with sufficient information for cases of *easy, a priori logical reasoning.*

that case, (EP*) allows that any proposition can be epistemically possible for such agents. Of course, this will have the consequence that agents that purport to believe p but not q, even though q follows from p by easy reasoning will be described as having no cognitive resources available for easy reasoning. *Had* we aimed for a model of human psychology, this would be implausible. Normal people often forget a relevant piece of information, make a mistake or get distracted in a given chain of reasoning. But for purposes of this project, I will abstract away from these aspects of non-ideal reasoning and focus exclusively on the aspects that pertain to limited cognitive resources. Obviously, this is an idealization. I will be able to model agents with limited cognitives resources, but not agents that make mistakes in their reasoning.

Given this, suppose we have a well-defined space of non-ideal scenarios, and suppose we want to use these scenarios to give a world involving analysis of a non-trivial notion of epistemic possibility that applies to minimally rational agents. Return to the example from above, and suppose a minimally rational agent a knows the proposition p that there are 2 apples in that basket and 3 apples in this basket. Then p is epistemically necessary for a. Since a can easily infer the proposition q that there are 5 apples from the proposition p, a should rationally accept q when she accepts p. We capture this normative element by saying that if p is epistemically necessary for a, then so is q. On the other hand, suppose a knows the proposition p_1 that there are 298 apples in each of 783 baskets. Then p_1 is epistemically necessary for a. But we can imagine that a cannot easily infer the proposition q_1 that there are 233,334 apples from the proposition p_1. Then a is rationally excused, on the current minimal conception of rationality, for not accepting q_1. We capture this normative element by saying that even though p_1 is epistemically necessary for a, q_1 may remain epistemically possible for a. To give a world involving analysis of this non-trivial notion of epistemic possibility, we then need to ensure that if p and

p_1 are true at each scenario that remains epistemically possible for a, then q is also true at each such scenario, while q_1 is false at some such scenario. But if so, this suggests that we need a construction of non-ideal scenarios that obey certain constraints and where not "anything goes". If everything went, we could not be assured that q is true at each epistemically possible scenario for a whenever p is true at each such scenario.

In this sense, *if* we can give a successful construction of non-trivial epistemic space, we might use it to develop an interesting world involving analysis of a non-trivial notion of epistemic possibility that applies to minimally rational agents. By having a space of scenarios that appropriately reflect what should remain epistemically possible for ordinary reasoners, the hope is that this space can potentially play a fundamental role in characterizing the modal borders within which we can model ordinary, yet non-trivial reasoning, inquiry and deliberation. Moreover, we could also use such a world involving analysis to shed light on the recent flow of papers that deal with various aspects of the non-trivial notion of epistemic possibility. For instance, we might use scenarios in non-trivial epistemic space to participate in the debates over "the" correct semantic analysis of expressions that involve epistemic modals such as 'might' and 'must'.[35] We might propose variations of the standard world involving contextualist or relativist treatments of epistemic modals, which can handle utterances of "It might be that not-s", where s is a mathematical truth. But in general, I trust that there is an independent interest in understanding epistemic possibility and necessity in their own right. Following Huemer, though epistemic possibility has not nearly received as much attention as metaphysical and logical possibility in the philosophical literature, it is arguably "the kind of possibility most often invoked in ordinary life."[36]

[35] See, for instance, Egan et al. (2005) and Egan (2007).

[36] Huemer (2007): p. 119.

Given this, I hope there is motivation for the content and rationality desiderata, and derivatively prima facie motivation for investigating constructions of non-trivial epistemic spaces. After having investigated Extreme Epistemic Space in chapter 2, I will argue that we have more than prima facie motivation for investigating non-trivial epistemic spaces. More specifically, I will motivate the following two claims. First, if we want to make non-trivial inferences from what obtains and does not obtain throughout a class of scenarios to the contents of the epistemic states of moderately ideal agents, then we need non-ideal scenarios to obey substantive constraints. Second, if we want to make non-trivial inferences from what obtains and does not obtain throughout a class of scenarios to what is epistemically possible and necessary for minimally rational agents, then we need non-ideal scenarios to obey substantive constraints. In both cases, this involves investigating non-trivial epistemic spaces.

Although both the content and rationality desiderata rely on a notion of non-trivial reasoning that tracks what can be established by easy or obvious reasoning, it is worth stressing that the content desideratum does not say anything about what agents should rationally believe. Rather, it says something about the contents of thoughts and beliefs that stand in non-trivial inferential relations to each other. The claim that the contents of thoughts and beliefs are determined partly by what can be established by easy or obvious reasoning is not in itself a normative claim. And one may well accept this claim while maintaining that there is more to minimal rationality than what can be established by easy or obvious reasoning. As such, a notion of hyperintensional content need not come with a built-in normative component of some sort.

Yet, as we shall see in later chapters, since both the content and rationality desiderata rely on a notion of easy or obvious reasoning, the two desiderata go together. This is already reflected in the examples above. As illustrated, to satisfy both the content and rationality desiderata, non-ideal scenarios need to verify the proposition q that there are 5 apples whenever they verify the proposition p that there are 2 apples in that basket and 3 apples in this basket. As such, if we can formulate a constraint on non-ideal scenarios that ensures that q is true at a scenario w whenever p is true at w, we can satisfy both the content and rationality desiderata. And if we cannot formulate such a constraint, we will fail to satisfy both desiderata. In this sense, the content and rationality desiderata go together. This also means that whenever the contents of p and q are the same—i.e. whenever p and q are true at the same scenarios—then any agent that believes p thereby also believes q and thus counts as a minimally rational agent in the relevant sense.

But although the content and rationality desiderata go together, it is clear that they are distinct desiderata. To illustrate this with an example, consider again the proposition p_1 that there are 298 apples in each of 783 baskets and the proposition q_1 that there are 233,334 apples. Although q_1 does not follow from p_1 by easy or obvious reasoning, suppose we say that q_1 nonetheless follows from p_1 by *feasible* reasoning. Suppose also that we have a good grip on the notion of feasible reasoning, and that we require that an agent should rationally believe a proposition p if the agent can come to believe p by feasible reasoning. We could then define a minimally rational agent as an agent that never violates this requirement, and let the rationality desideratum apply to such agents. To satisfy the rationality desideratum, we would then need a constraint on non-ideal scenarios that says that q_1 is true at a scenario w whenever p_1 is true at w. Yet, we might well maintain that the contents of p_1 and q_1 are different, in which case we would need non-ideal scenarios at which p_1 is true but q_1 false to satisfy the content desideratum. In such cases, the content and rationality desiderata come apart. So although the two desiderata go together throughout subsequent constructions of non-ideal epistemic space, they are distinct desiderata—just as the notions of content and rationality are distinct notions.

In investigating the prospects of constructing non-ideal epistemic spaces that can help us satisfy the content and rationality desiderata, the task of making (Epi-Pos*) and (Epi-Nec*) plausible for the broad class of non-ideal agents constitutes an important first step. If we cannot ensure that complex a priori falsehoods may remain epistemically possible for non-ideal agents, we cannot use the corresponding space to define non-ideal epistemic intensions that can play a role in analyzing non-trivial hyperintensional content. And neither can we use the corresponding space to give a world involving analysis of a non-trivial notion of epistemic possibility that applies to minimally rational agents. But if we can ensure that (Epi-Pos*) and (Epi-Nec*) are plausible principles for non-ideal agents, we can then evaluate whether the construction is suitable for satisfying the content and rationality desiderata.

1.6 Overview

The current project is essentially exploratory. Its main aim is to explore various constructions of non-ideal epistemic space and the pros and cons that they have—in particular, to explore and evaluate constructions of non-trivial epistemic spaces that impose substantive constraints on non-ideal scenarios or impossible worlds in a broad sense.

To anticipate my conclusions in a rough and simplified form: If the criterion of success is measured in terms of the content and rationality desiderata, then successful constructions of non-trivial epistemic spaces are hard, if not impossible to find. As we shall see, the general problem is that constructions of non-trivial epistemic spaces either pull too much towards spaces like Ideal Epistemic Space or too much towards spaces like Extreme Epistemic Space. If we want to avoid logical omniscience, the structure of a space like Ideal Epistemic Space is of little use. If we want to satisfy the content and rationality desiderata, the nearly trivial structure of a space like Extreme Epistemic Space is of little use.

More generally, if we want a construction of a non-ideal epistemic space that can safely navigate between the Charybdis of logical omniscience and the Scylla of "anything goes", our options are severely limited.[37] If you think, as I did, that there should be a wide spectrum of intermediate epistemic spaces to investigate, you will be surprised to find that there is not.

Before digging into the details, here is a brief chapter outline:

In chapter 2, I provide a construction of Extreme Epistemic Space. Effectively, Extreme Epistemic Space constitutes the diametric opposite of Ideal Epistemic Space, and it ensures that the (Epi-Pos*) and (Epi-Nec*) principles are plausible even for extremely non-ideal agents. Though we can potentially avoid all hyperintensional problems in Extreme Epistemic Space, I argue that we have good reason to investigate constructions that impose substantive constraints on non-ideal scenarios. This will lead us to constructions of non-trivial epistemic spaces.

In chapter 3, I lay out the general background structure for constructions of non-trivial epistemic spaces. By giving a test case interpretation of the notion of non-ideal deep epistemic possibility in terms of a notion of provability in n steps, I will set up the structure for stratified constructions of epistemic spaces. Roughly, the idea is to have a sequence of notions of non-ideal deep epistemic possibility that is associated with a corresponding sequence of non-

[37] Cf. Chalmers (forthcoming): p. 49:

> Perhaps the biggest open problem in the study of non-ideal epistemic space is that of finding a construction of non-ideal scenarios that avoids the Scylla of "anything goes" and the Charybdis of logical omniscience.

ideal epistemic spaces, the different end points of which are spaces much like Extreme Epistemic Space and spaces much like Ideal Epistemic Space. The hope is that we can find a particular space in this sequence that can help us satisfy the content and rationality desiderata.

In chapter 4, I investigate three intended constructions of non-trivial epistemic space. Though the first model, the Single Disprovability Model, can satisfy the core formal idea behind constructions of non-trivial epistemic spaces, I will argue that scenarios in the model remain too unconstrained to make serious progress on the content and rationality desiderata. I then investigate two alternative constructions—the Joint Disprovability Model and Jago's model in *Logical Information and Epistemic Space*—that attempt to impose further constraints on non-ideal scenarios.[38] Both models will have the undesired consequence that agents characterized by these models turn out to be logically omniscient, and that hardly any non-ideal scenarios survive in the corresponding epistemic spaces.

In chapter 5, I isolate the root of the problem that constructions of non-trivial epistemic spaces have. Roughly, I will show that attempts to eliminate extremely unconstrained non-ideal scenarios from non-ideal epistemic space have the undesired result that *all* non-ideal scenarios are eliminated from the corresponding space. Alternatively, as soon as we attempt to avoid the Scylla of "anything goes", we run into the Charybdis of logical omniscience. To avoid logical omniscience, this will leave us with two kinds of alternatives that I call Intermediate Models and Partial Models. I argue that both sets of alternatives fail to make progress on the content and rationality desiderata.

In chapter 6, I conclude and return to some of the issues discussed in this chapter.

[38] Jago (2009a).

Chapter 2

Extreme Epistemic Space

If we want to avoid logical omniscience, we need non-ideal scenarios in epistemic space. In this chapter, I investigate a maximally liberal model of non-ideal epistemic space that I call *Extreme Epistemic Space*. In contrast to Ideal Epistemic Space, we will be able to use scenarios in Extreme Epistemic Space to capture facts about epistemic possibility for all kinds of non-ideal agents.

First, however, I want to introduce a number of key elements that will play a central role for the basic framework and analyses of this project. Second, I lay down the properties that Extreme Epistemic Space needs to have to ensure that we can model extremely non-ideal agents. Third, I evaluate Extreme Epistemic Space. For the purpose of modeling the broad class of agents that are not extremely non-ideal, I argue that we have good reason to investigate models that have a less trivial structure than Extreme Epistemic Space. (In the appendix in section 2.6, I discuss a specification of a "scenario-making" language that allows us to give up two substantial assumptions that I introduce in this chapter and invoke for the general project.)

By the end of this chapter, we will be ready to investigate non-trivial epistemic spaces whose main job is to capture facts about epistemic possibility for moderately ideal agents. Such spaces, as we shall see, will have to navigate between Ideal Epistemic Space and Extreme Epistemic Space.

2.1 Acceptance and Rejection

What are the objects of epistemic possibility? For the motivational part, I have worked with propositions as a placeholder notion for the objects of epistemic possibility. But for much of the constructive work that I want to do in this project, it is useful to have some more definite entities to work with. The question is then which entities we should invoke.

One of the declared potential applications of non-ideal epistemic space is to motivate a notion of propositional content that although unstructured is as fine-grained as structured propositional content—and in particular, to make sense of a notion of Fregean propositions. So I cannot easily assume Fregean propositions from the outset. And at the same time it hardly makes sense to presuppose any of the other standard notions of propositions, even if they or their sufficiently enriched cousins could capture the kinds of cases that I discussed in the introduction. We could in principle think of propositions as primitive abstract entities that at least initially can capture arbitrarily fine-grained cognitive and epistemological differences in reasoning. But again, it is useful to adopt some more definite entities for many of my purposes.

More generally, there is also independent reason to circumvent the notion of propositions. The debates between Russellian and Fregean views, and between structured and unstructured views of propositions are complicated and multi-dimensional. So to avoid entangling the intuitive picture and the underlying world involving analysis of epistemic possibility with these debates, there is reason to bypass propositions altogether. For the main objectives of this project, the details of the extensive literature on propositions are mainly a distraction.

So instead I will let the objects of epistemic possibility be *sentences* and analyze what it means for a sentence to be epistemically possible and necessary in terms of the notions of *acceptance* and *rejection.* In the following sections, I will characterize these notions more precisely.

I make the following assumptions about acceptance and rejection. First, acceptance and rejection are to be understood as exclusive, though not exhaustive notions. Following Priest,

> acceptance and rejection do appear to be incompatible. One can certainly believe something and believe its negation. One might even argue that one can believe something and not believe it, though this is much more dubious. But it seems difficult to argue that one might both believe something and *refuse* to believe it. Characteristically, the behaviour patterns that go with doing X and refusing to do X cannot be displayed simultaneously.[1]

So while no agent can simultaneously accept and reject a sentence, the agent may well neither accept nor reject it. Second, acceptance and rejection are to be understood as dispositional notions: Roughly, as dispositions to occurrently accept and to occurrently reject. Though I will say more below about what it might mean to accept a sentence—more precisely in section 2.1.2—we can for now think of acceptance in terms of dispositions to assert or to assent to a sentence. Third, acceptance and rejection are to be understood as notions of *sincere* acceptance and rejection. Roughly, this means that agents that accept and reject various sentences are attentive, mentally well-functioning, and not under the influence of various drugs. Finally, I take it that acceptance and rejection of sentences can display a dynamic structure that reveals certain inferential patterns and regularities. For instance, if I accept 'If it rains in Canberra, the streets are wet in Canberra', I also typically accept 'The streets are wet in Canberra' should I come to accept 'It rains in Canberra'.

[1] Priest (2006b): pp. 98-99; refer also to Priest (1999): pp. 113-115 and Priest (2006a): pp. 109-110. Restall (2005a) refers to states in which a sentence is both accepted and rejected as 'self-defeating' states; see also Beall and Restall (2000), Humberstone (2000), Rumfitt (2000), and Smiley (1996).

For the initial construction of Extreme Epistemic Space, which is specifically targeted at extremely non-ideal agents, I will take both (occurrent) acceptance and (occurrent) rejection as primitive notions. This is motivated by the intuition that extremely non-ideal agents are the kinds of agents that may simultaneously accept sentence A and accept sentence $\neg A$. When I shift the focus to moderately ideal agents in chapter 3 and onwards, I will define 're-jection of A' as 'acceptance of $\neg A$' and only take the notion of acceptance as primitive.[2] This is motivated by the intuition that moderately ideal agents are the kinds of agents that never accept contradictions—*pace* dialetheism where it might be rational to believe certain contradictions.

2.1.1 Acceptance and Sentences

What is the language(s) in which agents accept and reject sentences? For the constructions that I will investigate in this project, the existence of multiple agent languages and of context-dependence causes problems. As mentioned in the introduction, I will focus on linguistic or ersatz constructions of scenarios as sets of sentences in some "scenario-making language". To analyze epistemic possibility and necessity, we will want to evaluate sentences for truth and falsity at scenarios constructed in this scenario-making language.

To illustrate the problems that arise from multiple agent languages and context-dependence, assume that we identify the objects of acceptance and rejection with (declarative) sentence types of multiple natural languages. To a first approximation, we can then say that an agent a accepts (rejects) a sentence type A just in case a is disposed to occurrently accept (reject) a token of A. If A is a type of a language that a does not understand, we can say that a neither accepts nor rejects A.

[2] If one has a better grip on the notion of rejection, one can of course also take rejection as primitive and let acceptance be the defined notion.

Given this, the key question is now which language the scenario-making language should be. For my purposes, there seems to be two options. On the first option, we construct scenarios in a single language $\mathcal{L}^+$—$\mathcal{L}^+$ might be a natural language like English, but it need not. In order to evaluate sentence types of another given language $\mathcal{L}_i$ at scenarios constructed in $\mathcal{L}^+$, we invoke translation relations between sentence types in $\mathcal{L}^+$ and sentence types in $\mathcal{L}_i$. Roughly, the job of the translation relation is to make sure that sentence types of arbitrary languages are suitable for being evaluated at scenarios constructed as sets of sentence types in a single language $\mathcal{L}^+$ that may be different from all these other languages. However, since an appeal to a translation relation raises a number of further difficult issues, I will postpone discussion of translation-involving constructions of scenarios to the appendix of this chapter.

On the second option, we dispense with a translation relation and aim to construct scenarios by using a combination of all natural languages simultaneously. To briefly illustrate and discuss this second option, assume that we let the scenario-making language correspond to the class $\mathcal{L}_U$ of (possible) sentence types of all natural languages. Scenarios will then correspond to sets of sentence types of all natural languages. To evaluate sentence types of a given natural language $\mathcal{L}_i$ at scenarios constructed in $\mathcal{L}_U$, a sentence type A of $\mathcal{L}_i$ will be true at a scenario if that scenario contains relevant sentence types of $\mathcal{L}_i$ that stand in the appropriate "true at" relation to A—for instance, the scenario might already contain the relevant sentence type A of $\mathcal{L}_i$. Since we are not appealing to a translation relation for the evaluation of sentence types at scenarios, any sentences in languages other than $\mathcal{L}_i$ will be irrelevant for the evaluation of A at scenarios constructed in $\mathcal{L}_U$.

The scenarios constructed in $\mathcal{L}_U$ will be of little use for many applications of non-ideal epistemic space. In particular, they will be of little use for defining non-ideal epistemic intensions that we can use to model *contents*. For if agents

a and b accept and reject types of different languages, there will always be arbitrarily many distinctions among scenarios constructed in $\mathcal{L}_U$ that a but not b can make, and that b but not a can make. Derivatively, the classes of scenarios that remain epistemically possible for agents that accept and reject types of different languages will always be different. And correspondingly, the non-ideal epistemic intensions that we can define over these scenarios will always be different for agents that speak different languages. Yet to model the *contents* of the epistemic states of arbitrary agents, this general language-relativity of non-ideal epistemic intensions is unwanted.

Even if we restrict the construction of scenarios to a single natural language, problems of context-dependence still arise. Suppose we construct scenarios from sentence types in a context-dependent language like English. Consider then the context-dependent type 'This is blue'. Clearly, many agents are simultaneously disposed to occurrently accept—think assert or assent to—some tokens of 'This is blue' and to occurrently reject some tokens of 'This is blue'. But on the definition above, such agents are then said both to accept and reject the type 'This is blue', which is impossible in the current framework. Since this pattern generalizes to many context-dependent types, context-dependent phenomena in natural languages then pose a problem for analyses that identify the objects of acceptance and rejection with types of natural languages.

We might attempt to avoid the problems concerning context-dependence by identifying the objects of acceptance and rejection with sentence *tokens* of a natural language, and by constructing scenarios as sets of these tokens. But as the construction of scenarios in $\mathcal{L}_U$, the corresponding construction of scenarios as sets of sentence tokens remains of little value for many applications of non-ideal epistemic space. In particular, since sentence tokens are tied to specific agents, the resulting epistemic space will be completely agent-relative. But for the purposes of capturing facts about epistemic possibility for classes of non-ideal agents, and for specifying non-ideal epistemic intensions that allow us to model general features of non-ideal content and reasoning, this agent-relativity is unwanted.

One might attempt to handle both the multiple-language problem and the context-dependence problem by again appealing to a translation relation. For instance, we might construct scenarios in a single context-independent language $\mathcal{L}^{+}$—since context-dependent phenomena are abundant in natural languages like English, $\mathcal{L}^{+}$ cannot be a natural language. In order to evaluate sentence tokens of arbitrary natural languages for truth and falsity at scenarios constructed in $\mathcal{L}^{+}$, we invoke translation relations between sentence types in $\mathcal{L}^{+}$ and sentence tokens. But again, since the appeal to a translation relation generates a number of further complications, I will postpone discussion of this kind of construction to the appendix of this chapter.

To handle the multiple-language problem and the context-dependence problem in a more straightforward manner, I will make two strong assumptions—in the appendix in section 2.6 below, I then discuss a way in which we can give up these assumptions. First, I will assume that we want to model only agents that accept and reject sentences in a single language $\mathcal{L}$. Call this the *language assumption*. Second, I will assume that there is no context-dependence in $\mathcal{L}$, or better, that agents are never disposed to simultaneously accept or reject tokens of the same type in $\mathcal{L}$. Call this the *context assumption*. Clearly, the language and context assumptions are very strong. The main motivation for these assumptions lies in the problems above, and as I will notice below, they are not unreasonable for my main purposes in this project.

Given the language and context assumptions, the question is then which kind of language $\mathcal{L}$ is. Though $\mathcal{L}$ cannot be a natural language like English because of context-dependent phenomena, we can let $\mathcal{L}$ be a regimented version of English—sufficiently enriched with mathematical, logical and scientific expressions—that I will call *English**.[3] The rough idea behind English* is this: Whereas speakers of English are often disposed both to assert and to deny tokens of a given sentence type to convey different pieces of information, speakers of English* can always use tokens of orthographically distinct types to convey similar pieces of information. For instance, speakers of English are often disposed both to assert and to deny tokens of 'The banks are beautiful' to express their beliefs in the beauty of river banks and their disbeliefs in the beauty of financial institutes. But speakers of English* can always use tokens of orthographically distinct types such as 'The banks$_1$ are beautiful' and 'The banks$_2$ are beautiful' to express distinct beliefs.

To make this rough idea behind English* more precise, we can invoke Chalmers' distinction between *epistemically variant* and *epistemically invariant expressions*.[4] To illustrate the idea behind the distinction, consider the name 'Neptune'. Suppose Leverrier uses 'Neptune' as a name for whatever planet perturbs the orbit of Uranus. For Leverrier, the inference from 'Neptune is beautiful' to 'The planet perturbing the orbit of Uranus is beautiful' can then be justified a priori. But for later speakers like me that have picked up the name 'Neptune' through some empirical channel, this inference cannot be justified a priori. When a term like 'Neptune' supports differences in apriority in this way, the term is epistemically variant. Likewise, a term such as 'tall' is epistemically variant because of the different standards of tallness that speakers may associate with the term. On one occasion, a speaker may use 'tall' to pick out all people above 2 meters, but yet on another occasion to pick out all people above 1.8 meters. Correspondingly, the inference from

[3] A similar observation is made by Jeffrey (1983). Much like I do, Jeffrey stipulates that the relevant agent language—in which he also aims to construct his ersatz worlds or complete consistent novels—must be "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.)

[4] Cf. Chalmers (2003): p. 50, and Chalmers (forthcoming): p. 9.

'If someone is tall, then she is happy' to 'If someone is above 2 meters, then she is happy' may be a priori justifiable on some occasions of use, but not on others.

More generally, when an expression such as 'Neptune' and 'tall' "supports potential differences in apriority among fully competent users in this way", we can say that the expression is epistemically variant; if not, the expression is epistemically invariant.[5] Given this, we can now aim to specify English* as a regimented version of English in which all (epistemically) variant expressions in English are replaced by invariant expressions. To obtain such a class of invariant expressions, we can regiment English (at least) in the following ways. First, ambiguous terms such as 'bank' are replaced in English* by distinct terms 'bank$_1$' and 'bank$_2$' reflecting the disambiguations of 'bank'. Second, context-dependent terms such as 'tall' and 'bald' are replaced in English* by arbitrarily many terms 'tall$_n$' and 'bald$_n$' reflecting all the different standards for tallness and baldness. Third, demonstrative terms such as 'that' are replaced in English* by arbitrarily many terms 'that$_n$', each of which reflects, roughly, what an agent using the relevant demonstrative intends to refer to on a given occasion of use. Proper names such as 'Neptune' are replaced in English* by arbitrarily many names 'Neptune$_n$', each of which reflects, roughly, the mode of presentation that an agent associates with the relevant name on a given occasion of use.[6]

If we keep applying this process of regimentation, we will plausibly be left with a version of English in which all epistemically variant types of English are replaced by corresponding epistemically invariant types. This is the language English*. Plausibly, certain indexical terms like 'I' and various mathematical and logical terms such as 'two' are epistemically invariant—at least when used

[5] Chalmers (forthcoming): p. 9.

[6] For details on the epistemic variance of demonstratives and names, see Chalmers (2002a): pp. 173-174.

by fully competent speakers of English. If so, then such terms can survive in English*. Accordingly, since all sentence types in English* are composed of epistemically invariant expressions, all sentence types in English* are epistemically invariant. Intuitively, no two sentence types in English* will support differences in apriority among fully competent speakers of English*.

Clearly, English* is not the kind of language that beings like us could easily speak. Though we can easily imagine speaking a version of English that contains two words 'bank$_1$' and 'bank$_2$' instead of the single ambiguous 'bank', it is much harder to imagine speaking a version of English that contains a potential infinite class of invariant expressions that have replaced all variant expressions in English that involve names and demonstratives.[7] So although the idealizations involved in the specification of English* are obvious, I trust that we have a good grasp of the nature of a language like English*.

Generally, however, it is worth making the following observation: The current idealizations do not entail a relevant idealization of the cognitive capacities of agents. Of course, given the complexity of English*, we might be idealizing an agent's capacities for learning and understanding a language. But the fact that agents are speakers of English* can never in itself imply that agents are mathematically or logically omniscient. For instance, though an agent accepts the sentences A and $(A \rightarrow B)$ in English*, we are of course not forced to say that the agent also accepts B *because* the agent is a speaker of English*. So for the kinds of non-ideal agents that I am interested in, the idealizations do not seem unreasonable. That is, the main focus is on resource-bounded agents that can engage in limited, but non-trivial (logical) reasoning, and for the purposes of analyzing epistemic possibility and necessity in light of these aspects of non-

[7] Given that we assume a finite number of finite agents, each of which in principle accepts and rejects at most a countable number of distinct sentence types, we can say that English* will at most contain a countable number of invariant expression types, and also that all sentence types in English* will be of finite length.

ideal reasoning, the idealizations above do not seem unreasonable. And more generally, for investigating the *prospects* of constructing non-ideal epistemic spaces that can satisfy the content and rationality desiderata—and hence play the role of non-trivial epistemic space—the idealizations above do not affect the main conclusion that I want to establish.

So given all this, we can identify the agent language $\mathcal{L}$ with English*. So henceforth—except for section 2.6 below—we should understand all unqualified talk of sentences to refer to sentence types in $\mathcal{L}$. Since all epistemic variance is eliminated from $\mathcal{L}$, we can then plausibly say that fully competent speakers of $\mathcal{L}$ are never disposed to simultaneously accept and reject tokens of the same type in $\mathcal{L}$. So given the language and context assumptions, $\mathcal{L}$ is hence a suitable language for this project that allows us to identify the "scenario-making" language with the language $\mathcal{L}$ in future constructions of scenarios and non-ideal epistemic spaces.

2.1.2 Acceptance and Thoughts

To a first approximation, we can now say that an agent a *accepts* (rejects) a sentence type A in $\mathcal{L}$ just in case a is disposed to occurrently accept (reject) a token of A. What does it mean to occurrently accept and to occurrently reject a token of a sentence type in $\mathcal{L}$? As mentioned, I will take 'occurrent acceptance' and 'occurrent rejection'—and in a sense then also 'acceptance' and 'rejection'—as primitive notions. The motivation is here that multiple interpretations of these notions are compatible with the basic framework, and as such, that multiple refinements of the general definition are possible.[8]

For many of my purposes, however, a useful heuristics for understanding acceptance centers around a Chalmersian conception of *thoughts*:

> A thought is understood here as a token mental state, and in particular as a

[8] For a survey of some of the many interpretations of 'acceptance', see Pascal (1998).

> sort of occurrent propositional attitude: roughly, an entertaining of a content. The idea is that this is the sort of propositional attitude that is generally expressed by utterances of assertive sentences. Such utterances typically express occurrent beliefs, but they do not always express occurrent beliefs, as subjects do not always believe what they say. Even in these cases, however, the subject *entertains* the relevant content: a thought is an entertaining of this sort. Like beliefs, thoughts are assessable for truth. Thoughts can come to be *accepted*, yielding beliefs, and thoughts can come to be *justified*, often yielding knowledge. When an utterance expresses a thought, the truth-values of the utterance and the thought always coincide.[9]

Intuitively, typically we know what we are thinking, and typically we are in a position to find a sentence that we can use to express the contents of our thoughts. Though we need not invoke any strong ‘internal’ relation between thought and utterance content, it is useful to invoke a general relation of expression, according to which every sincere assertive utterance expresses a thought.

Given the notion of thoughts, we can now define ‘occurrent acceptance’ in terms of accepting the content of a corresponding thought, where we can take what it means to accept the content of a thought as an intuitive primitive—roughly, an endorsement of a content as true, where the kind of endorsement is the one that is characteristic of belief. We can then say that an agent a *occurrently accepts* (rejects) a token of a type A in $\mathcal{L}$ just in case a accepts (rejects) the content of a thought that a expresses by (sincerely) uttering A. We can then refine the definition above and say that an agent a accepts a sentence type A in $\mathcal{L}$ just in case a accepts the content of a thought that a is disposed to express by uttering A.

We can now also introduce a notion of *justified acceptance* that mirrors knowledge: An agent a *justifiably accepts* a type A in $\mathcal{L}$ just in case a is

[9] Chalmers (2004): p. 96.

justified in accepting the content of a thought that she is disposed to express by uttering A, where justified acceptance of a thought often yields knowledge. As such, we can say that mere acceptance of a thought yields belief, and hence that mere acceptance of a type in $\mathcal{L}$ mirrors belief.[10]

Further, we can also invoke an intuitive notion of thought or belief content that directly tracks the fine-grained cognitive and epistemological differences that the current notions of epistemic possibility and necessity require. For instance, we can say that most agents accept 'Water is water' and '$2 + 2 = 4$' because the thoughts typically expressed by tokens of these sentence types can be justified through a cognitively trivial chain of a priori reasoning.[11] In contrast, there is no chain of a priori reasoning that begins with the thought typically expressed by a token of 'Water is H_2O' and leads to an acceptance of that thought. Intuitively, mere reflection on the thought expressed by a token of 'Water is H_2O' will not reveal the truth of that very thought. Further, though the thought typically expressed by a token of 'There are no integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$' can be justified a priori, it need not be the case that non-ideal agents can emulate this highly non-trivial and cognitively demanding chain of a priori reasoning. Intuitively, mere non-ideal reflection on the thought expressed by a token of 'There are no integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$' will not reveal the truth of that very thought.

Finally, we can use the notion of thoughts to relate familiar logical inferences between sentences to corresponding inferences in thought. Following Chalmers, we can plausibly hold that

> [. . .] the thoughts of a given thinker can stand in [. . .] relations of negation, conjunction, and disjunction to each other: so one thought can be formed

[10] As mentioned in the introduction, since the relevant insights of this project apply both to notions of doxastic and epistemic possibility, I will largely ignore the distinction between acceptance and justified acceptance in the following.

[11] For purposes of illustration here and elsewhere, I always assume that the relevant English sentence types are also sentence types in $\mathcal{L}$.

> by another by an operation of negation, or from another two thoughts by operations of conjunction or disjunction.[12]

Derivatively, this can provide us with an explanation of why many agents that accept types A and B also typically accept the conjunction $(A \wedge B)$.

So I trust that the notion of thoughts provides a useful way of understanding the notions of acceptance and rejection, and we can take this heuristic characterization to be operative in the background. But alternative characterizations are available. For instance, we might characterize acceptance in broadly behavioristic terms and say that an agent a occurrently accepts a sentence type A in $\mathcal{L}$ just in case a assents to or asserts A (whether in private or in public). Or alternatively, we can say that an agent a occurrently accepts a sentence type A in $\mathcal{L}$ just in case a answers 'yes' to the question "Is it true that A?".[13] Derivatively, we could use these definitions to refine the general definition of acceptance and rejection above. In this sense, the basic framework is not committed to any particular interpretation of these primitive notions.

2.1.3 Acceptance and Epistemic Possibility

Given a characterization of the notions of acceptance and rejection of sentence types in $\mathcal{L}$, we can now restate the (EP*) and (EN*) analyses as follows, where unqualified talk of sentences, as mentioned, should always be understood as involving talk about sentence types in $\mathcal{L}$:

(**EP**) A sentence A is epistemically possible for an agent a iff a cannot easily come to reject A by a priori reasoning from what a already accepts and rejects.

(**EN**) A sentence A is epistemically necessary for an agent a iff a can easily come to accept A by a priori reasoning from what

[12] Chalmers (forthcoming): p. 9.

[13] For further details and discussion, see for instance Horwich (1998) and Speaks (2006).

a already accepts and rejects.

Since I initially take both (occurrent) acceptance and (occurrent) rejection as primitive, I hence initially take both epistemic possibility and necessity as primitive. Though no sentence A can be epistemically possible and epistemically impossible for any agent, both A and $\neg A$ can be epistemically necessary for extremely non-ideal agents.

When I shift the focus to moderately ideal agents, we can define epistemic possibility in terms of acceptance: A is epistemically possible for an agent a if and only if a cannot easily come to accept $\neg A$ by a priori reasoning from what a already accepts. Derivatively, we can recover the standard interdefinability of modal notions and define epistemic possibility in terms of epistemic necessity: A is epistemically possible for a if and only if $\neg A$ is not epistemically necessary for a. The extra structure that emerges from defining epistemic possibility in terms of epistemic necessity will be useful for later constructions of non-ideal epistemic space. But we should notice that all subsequent models of epistemic space will enable us to model agents for whom arbitrary *explicit* contradictions of the form $(A \wedge \neg A)$ are epistemically necessary. In this sense, we will always be able to accommodate extremely non-ideal agents that accept contradictions.

In what follows, I will focus on a priori reasoning rather than armchair reasoning. Most people, I take it, allow that a piece of armchair reasoning can be justified by exploiting existing empirical information. For instance, a piece of armchair reasoning that leads from A to B may be justified by certain general inductive principles such as 'the future resembles the past'. Plausibly, the justification for such general inductive principles ultimately relies on some empirical information about the world—intuitively, an ideal agent with unbounded cognitive capacities for a priori reasoning could entertain the possibility that she lives in a world where the future does not resemble the past.

If so, then the justification for the inference from A to B itself relies on empirical information. Since my primary aim is to establish facts about epistemic possibility that apply to all agents, irrespective of what empirical information they have, if any at all, the current conception of a priori reasoning should be understood to exclude any such reliance on empirical information. We can put this by saying that the standard of justification associated with the kind of a priori reasoning that figures in (EP) and (EN) reflects the "conclusive standard associated with proof and analysis."[14] On this conception, if a sentence can be accepted by purely a priori reasoning, then the possibility that the sentence is false can be ruled out conclusively.

I understand the kind of *easy* a priori reasoning that figures in (EP) and (EN) to be the kind of reasoning that unfolds in relevantly short cognitive episodes for an agent. Roughly, we can think of these short cognitive episodes for an agent as falling within the specious present for the agent in question. Or roughly, if an a priori inference from A to B is easy for an agent a, then the cognitive episode that involves B and that succeeds in time the cognitive episode that involves A will be considered or experienced as the same cognitive episode by a. As such, the notion of easy reasoning is dynamic. But the dynamics should be understood at a very coarse level of grain that abstracts away from whatever happens in the dynamics in the short cognitive episodes for an agent. So whenever I say that an agent can easily, immediately or instantly come to accept or reject a sentence by a priori reasoning, I have this kind of coarse grained dynamic a priori reasoning in mind.

2.2 Extreme Epistemic Space

The stage is now set to investigate the first construction of non-ideal epistemic space. To avoid logical omniscience, we know that we need non-ideal

[14] Chalmers (2004): p. 99.

scenarios in epistemic space. The question is then *how* non-ideal such scenarios should be. If there are agents for whom "[n]o inferences, however obvious and useful, need be made from the beliefs", and for whom the "the belief set can include any and all inconsistencies", we indeed need extremely non-ideal scenarios in epistemic space.[15] Since such extremely non-ideal agents need not have any cognitive capacities available for instant a priori reasoning, every sentence, and in particular every sentence in the a priori domain may remain epistemically possible for such cognitively impaired creatures. If so, then for *any* sentence A, we need scenarios where A is false to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles even for extremely non-ideal agents:

(**Epi-Pos**) A is epistemically possible for an agent a iff there is a scenario w in W such that w is epistemically possible for a and such that A is true at w.

(**Epi-Nec**) A is epistemically necessary for an agent a iff for each scenario w in W such that w is epistemically possible for a, A is true at w.

To ground a space of scenarios that can play this role, we hence need a maximally liberal notion of deep epistemic possibility, according to which any sentence is deeply epistemically possible.

This maximally liberal notion of deep epistemic possibility naturally leads to a construction of *Extreme Epistemic Space.* Scenarios in Extreme Epistemic Space are akin to Priest's *open worlds.*[16] Just like open worlds, there will for any sentence A be a scenario w in Extreme Epistemic Space such that A is true at w. And just like open worlds, there will for any set of sentences $\{A_1, A_2, \ldots\}$ be a scenario w in Extreme Epistemic Space such that the truths

[15] Cherniak (1986): p. 6.

[16] Cf. Priest (2005).

at w are exactly $A_1, A_2, \ldots$. Strictly, we do not need the kinds of scenarios that will be permitted in Extreme Epistemic Space to model the maximally liberal notion of deep epistemic possibility. In chapters 4 and 5, I investigate alternative constructions of non-ideal epistemic space that can also do the job, but Extreme Epistemic Space constitutes a natural point of departure.

For the construction of Extreme Epistemic Space, I first identify a scenario w with an arbitrary set of sentence types in the agent language $\mathcal{L}$—that is, given the assumptions above, with arbitrary sets of possible sentence types in English*. Second, I define what it means for a sentence A to be true or false at a scenario as follows:

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

(**Falsity**) A sentence A is false at scenario 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 define two scenarios w and w' to be *equivalent* if and only if for all A, $A \in w$ if and only if $A \in w'$.

Given these simple definitions, we can immediately establish that scenarios obey the following two principles:[17]

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

(**Parsimony**) If scenarios w and w' are equivalent, then $w = w'$.

Because $A \in w$ or $A \notin w$, for each sentence A and scenario w, (Basic Maximality) follows trivially from (Truth) and (Falsity). By individuating scenarios extensionally in terms of the sentences that they contain, (Parsimony) also

[17] The (Parsimony) principle is taken from Chalmers (forthcoming).

follows immediately. Let W_E be the class of scenarios that satisfy these basic principles.

Intuitively, (Basic Maximality) ensures that scenarios 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 scenario w describes?" will be answered yes if $A \in w$, and no if $A \notin w$.[18] So whereas ideal scenarios are maximally specific, a priori consistent sets of sentences, non-ideal scenarios are maximally specific sets of sentences some of which may fail to be a priori consistent. In chapters 4 and 5, I will investigate models of non-ideal epistemic space in which sentences may be indeterminate in truth-value at scenarios. But to keep things as familiar and as simple as possible from the outset, (Basic Maximality) is desirable.

It is worth stressing that (Basic Maximality) in conjunction with (Truth) and (Falsity) just is an instance of the trivial principle that for all sentences A and sets Γ, either $A \in \Gamma$ or $A \notin \Gamma$. (Basic Maximality) contrasts with a more standard, linguistic ersatz formulation of 'maximality', according to which either A is true at w or $\neg A$ is true at w, for all sentences A and scenarios w. In chapter 4, I will adopt such a formulation of 'maximality' myself. Yet, since we want to use Extreme Epistemic Space to model agents for whom, say, both A and $\neg A$ are epistemically impossible, (Basic Maximality) is an appropriate principle for this initial construction.

(Parsimony) says that if two scenarios are equivalent descriptions of a maximally specific way things might be, then one of them is redundant for the purpose of giving a world involving analysis of epistemic possibility. Intuitively, if two scenarios w and w' verify exactly the same sentences, then w and w' are epistemically indistinguishable, and as such, only one of them is needed in epistemic space.

[18] Cf. Jeffrey (1983): pp. 208-209.

Given this simple construction of Extreme Epistemic Space, we can now define what it means for a scenario w to be epistemically possible for an agent:

> (**Epi-Pos-**w) A scenario w is epistemically possible for an agent a iff for all sentences A:
>
> (i) If A is epistemically necessary for a, then $A \in w$.
>
> (ii) If A is epistemically impossible for a, then $A \notin w$.

We can then identify W with W_E and use (Epi-Pos-w) to prove the essential (Epi-Pos) and (Epi-Nec) principles:

> (**Epi-Pos**) A is epistemically possible for an agent a iff there is a scenario w in W such that w is epistemically possible for a and such that A is true at w.
>
> (**Epi-Nec**) A is epistemically necessary for an agent a iff for each scenario w in W such that w is epistemically possible for a, A is true at w.

The proof of (Epi-Pos) and (Epi-Nec) is easy:

> For (Epi-Pos) left to right, assume A is epistemically possible for a. Then A is not epistemically impossible for a. By (Epi-Pos-w), clause (ii), then it is not the case that $A \notin w$ for each epistemically possible scenario for a. So there is a scenario w such that w is epistemically possible for a and such that $A \in w$. For (Epi-Pos) right to left, assume A is true at some epistemically possible scenario w for a. Assume, for reductio, that A is epistemically impossible for a. By (Epi-Pos-w), clause (ii), then $A \notin w$ for each epistemically possible scenario w for a. By assumption, however, there is an epistemically possible scenario w for a such that $A \in w$. Hence A

is not epistemically impossible for a. So A is epistemically possible for a. So (Epi-Pos) holds.

For (Epi-Nec) left to right, assume A is epistemically necessary for a. By (Epi-Pos-w), clause (i), then $A \in w$ for each epistemically possible scenario for a. For (Epi-Nec) right to left, assume A is true at each epistemically possible scenario w for a. Assume, for reductio, that A is not epistemically necessary for a. By (Epi-Pos-w), clause (i), then $A \notin w$ for some epistemically possible scenario for a. By assumption, however, all epistemically possible scenarios w for a are such that $A \in w$. So A is epistemically necessary for a. So (Epi-Nec) holds. □

It is worth noting that this proof of (Epi-Pos) and (Epi-Nec) will hold for all subsequent models of non-ideal epistemic space. In some of the later models, however, we can simplify the proof of (Epi-Pos). As mentioned above, when I shift the focus to moderately ideal agents, we can define rejection in terms of acceptance, and derivatively define epistemic possibility in terms of epistemic necessity. With an appropriate constraint on scenarios that says that A is true at w if and only if $\neg A$ is false at w, we can straightforwardly prove (Epi-Pos) and (Epi-Nec) using just clause (i) in (Epi-Pos-w).[19] But for the current picture, we need both clauses in (Epi-Pos-w).

2.3 The Liberty of Extreme Epistemic Space

Given the construction of Extreme Epistemic Space, we can unproblematically ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for non-ideal agents. For instance, let s be Fermat's Last Theorem. Since there are plenty of scenarios $w \in W_E$ such that $s \notin w$, s will be false at any w that does not

[19] For details, see footnote 6, page 106. For the Partial Models that I investigate in chapter 5, section 5.3, the proof of (Epi-Pos) and (Epi-Nec) will proceed as above.

contain s. If so, then s does not have to be true at each scenario that remains epistemically possible for any given agent. By (Epi-Nec), then s does not have to be epistemically necessary for any agent. So in contrast to Ideal Epistemic Space, we can now model agents that fail to accept mathematical truths such as Fermat's Last Theorem.

More generally, we can pick *any* sentence A and find a scenario $w \in W_E$ that falsifies A. In particular, for any logical truth A, there is a $w \in W_E$ such that $A \notin w$. So we can easily model agents for whom no logical truths, however obvious, are epistemically necessary. Since extremely non-ideal agents may fail to accept even the most obvious logical truths, we can hence unproblematically model these kinds of agents in Extreme Epistemic Space. Further, assume $(A \wedge B)$ is epistemically necessary for some extremely non-ideal agent a. Then $(A \wedge B)$ is true at all w that are epistemically possible for a. Yet, since there are plenty of scenarios $w \in W_E$ such that $(A \wedge B) \in w$ but $A \notin w$ and $B \notin w$, A and B need not be epistemically necessary for a even when $(A \wedge B)$ is. Hence a need not accept A and B even when a accepts $(A \wedge B)$. So we can easily model agents that fail to accept even the most simple a priori consequences of what they accept.

Accordingly, Extreme Epistemic Space can do the intended work: It gives us a space of scenarios that we can use to capture facts about epistemic possibility for extremely non-ideal agents. In fact, there is reason to believe that we can use Extreme Epistemic Space to avoid all hyperintensional problems. To illustrate, let the *extreme epistemic intension* of a sentence be a function from scenarios in W_E to a truth-value. For any two sentences A and B, there is a scenario $w \in W_E$ such that w verifies A and falsifies B. Derivatively, for any two sentences A and B, the values of the extreme epistemic intensions of A and B can come apart in Extreme Epistemic Space.

To avoid hyperintensional problems in the philosophy of language, we can attempt to understand semantic content in terms of extreme epistemic intensions. Since the extreme epistemic intensions of any two sentences need never coincide in truth-value, Frege type puzzles and belief ascriptions need no longer cause problems. For instance, consider any two mathematically or logically equivalent A and B. For some $w \in W_E$, we have $A \in w$ but $B \notin w$. So the extreme epistemic intension of 'Julie believes A' can be true while the extreme epistemic intension of 'Julie believes B' can be false. Derivatively, we can capture the linguistic intuition that it can be true to say that 'Julie believes A' but at the same time false to say that 'Julie believes B' for many mathematically and logically equivalent A and B. For counterpossibles, we can pick any (intensionally) impossible antecedent A and rest assured that the extreme epistemic intension of A is true at some $w \in W_E$. If we can define a relevant closeness or similarity measure on scenarios in W_E, we can then ensure that counterpossibles can differ in truth-value.[20]

To avoid hyperintensional problems in the philosophy of mind, we can attempt to represent the contents of thoughts and beliefs by extreme epistemic intensions. For instance, since the class of scenarios in W_E that verify '$2{+}2 = 4$' need not coincide with the class of scenarios that verify '$44 + 88 = 132$', the hyperintensions of the thoughts that $2 + 2 = 4$ and that $44 + 88 = 132$ need not coincide either. Derivatively, we can represent agents as believing that $2 + 2 = 4$ without thereby also representing these agents as believing that $44 + 88 = 132$. Since Extreme Epistemic Space allows us to draw maximally fine-grained distinctions among epistemic possibilities in this way, extreme epistemic intensions can allow us to model a maximally fine-grained notion of mental content.

To avoid hyperintensional problems in epistemology, we can attempt to use W_E as the underlying space of possibilities that figures in various world models

[20] See Nolan (1997) for details.

in epistemology. For instance, we can attempt to define the standard knowledge operator K from epistemic logic in terms of quantification over scenarios in W_E. Consider then any logical truth A. Since A is false at some $w' \in W_E$, $K(A)$ can be false at any w from which w' is epistemically accessible. So agents need not be modeled as knowing all logical truths. Further, even if $K(A)$ is true at some $w \in W_E$, then some B that is logically equivalent to A might well be false at some epistemically accessible scenario from w. Derivatively, $K(B)$ need not be true at w. So agents need not be modeled as knowing everything logically equivalent to what they already know.[21] Accordingly, if W_E is used as the underlying space of possibilities, worries about logical omniscience need no longer arise.

Obviously, these remarks are not meant to imply that the *mere* existence of a space of possibilities like Extreme Epistemic Space can give us satisfactory theories in the relevant areas of philosophy of language, mind, and epistemology. Rather, the observation is simply that we can *avoid* all hyperintensional problems in Extreme Epistemic Space. This feature, I take it, is the main benefit of Extreme Epistemic Space. Yet, as I will argue now, there are good reasons to look beyond Extreme Epistemic Space.

2.4 Limitations of Extreme Epistemic Space

We have seen that the maximal liberty of Extreme Epistemic Space is useful for modeling the epistemic states of extremely non-ideal agents. Insofar as extremely non-ideal agents have no cognitive capacities available for instant a priori reasoning, and insofar as any set of sentences can represent the epistemic states of such agents, Extreme Epistemic Space can do its job. And if our primary aim is to ensure that we can draw maximally fine-grained distinctions

[21] For discussions of such impossible world involving models for epistemic logic, see Fagin et al. (1995): pp. 357-362 and Wansing (1990).

among possibilities, Extreme Epistemic Space can do this job too. For any A and B, if a distinction between A-possibilities and B-possibilities is required for a particular theoretical purpose, Extreme Epistemic Space will deliver such a distinction.

Extreme Epistemic Space can do these jobs because of its "explosive anything goes" nature: For "the most absurd situations conceivable", there is a corresponding scenario in W_E.[22] Everything goes in Extreme Epistemic Space because there are no non-trivial structural and formal constraints on scenarios in W_E. Pick ten arbitrary sentences $A_1, A_2, \ldots, A_{10}$, collect them in a set Γ, and you have a scenario $w \in W_E$ that corresponds to Γ. So arbitrary blatant inconsistencies like '$0 = 1$', 'My bike is red and blue all over', and 'It rains and it does not rain' are among the sentences verified by a particular $w \in W_E$. And arbitrary joint inconsistencies like {'It is summer', 'It is not summer' } and {'It is summer and it is sunny', 'It is not summer', 'It is not sunny'} are among the sentences verified by a particular $w \in W_E$.

The extreme epistemic intensions that we can define in Extreme Epistemic Space can trivially deliver different truth-values for any pair of sentences A and B since they have no more structure than A and B themselves. For instance, suppose $(A \wedge B) \in w$, for some $w \in W_E$. Then the extreme epistemic intension of $(A \wedge B)$ is true at w. Suppose also that $\neg A \in w$ and $\neg B \in w$. Then the extreme epistemic intensions of $\neg A$ and $\neg B$ are also true at w. In this sense, the behavior of the extreme epistemic intension of $(A \wedge B)$ is completely detached from the behavior of the extreme epistemic intensions of A and B. Though 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 this structure is of less use for modeling the broad class of agents that are not extremely non-ideal. To see this, we can consider the content and

[22] Nolan (1997): p. 544.

rationality desiderata.

For the content desideratum, we want to use scenarios in non-ideal epistemic space to give a world involving account of a non-trivial notion of hyperintensional content. In particular, we want to use non-ideal epistemic intensions to represent the contents of the epistemic states of moderately ideal agents. Since the contents of such epistemic states can stand in non-trivial inferential relations to each other, we want to reflect these relations in the non-ideal epistemic intensions. Consider an inference from $(A \wedge B)$ to A. If anything, I take it, the chain of reasoning that proceeds from $(A \wedge B)$ to A is as easy and computationally feasible as it gets. To reflect such basic inferential relations among thoughts and sentences in the corresponding non-ideal epistemic intensions, we need to ensure that the non-ideal epistemic intension of A is true at scenario w whenever the non-ideal epistemic intension of $(A \wedge B)$ is true at w. Extreme Epistemic Space cannot do this job: Though the extreme epistemic intension of $(A \wedge B)$ is true at some $w \in W_E$, there is no guarantee that the extreme epistemic intension of A is also true at w. For all the construction of Extreme Epistemic Space says, it may well be that the extreme epistemic intension of A is false at w while the extreme epistemic intension of $(A \wedge B)$ is true at w. So extreme epistemic intensions cannot play the role that we need non-ideal epistemic intensions to play to satisfy the content desideratum. We can call this the *content problem.*

For the rationality desideratum, we want to use scenarios in non-ideal epistemic space to give a world involving analysis of a non-trivial notion of epistemic possibility that captures which sentences should remain epistemically possible for minimally rational agents. Consider a minimally rational agent a that accepts $(A \wedge B)$. Since a can easily infer A from $(A \wedge B)$, a should rationally accept A when she accepts $(A \wedge B)$. We want to capture this normative element in the (EN) analysis of epistemic necessity by saying that if $(A \wedge B)$

is epistemically necessary for a, then so is A. To analyze such situations in terms of scenarios, we need to ensure that if $(A \wedge B)$ is true at all epistemically possible scenarios for a, then A is also true at all such scenarios. Extreme Epistemic Space cannot do this job: Though $(A \wedge B)$ is true at each $w \in W_E$ that remains epistemically possible for a, there is no guarantee that A is true at w. For all the construction of Extreme Epistemic Space says, it may well be that $(A \wedge B)$ and $\neg A$ are true at each w that remains epistemically possible for a. So scenarios in Extreme Epistemic Space cannot play the role that scenarios need to play to satisfy the rationality desideratum. We can call this the *rationality problem*.

Accordingly, if we want to make non-trivial inferences from what obtains and does not obtain throughout a class of scenarios to the contents of the epistemic states of moderately ideal agents, scenarios need to obey certain substantive constraints that scenarios in Extreme Epistemic Space do not obey. And if we want to make non-trivial inferences from what obtains and does not obtain throughout a class of scenarios to what is epistemically possible for minimally rational agents, scenarios need to obey certain substantive constraints that scenarios in Extreme Epistemic Space do not obey. As exemplified, if $(A \wedge B)$ is true throughout a class of scenarios that remain epistemically possible for a moderately ideal agent, A and B should also be true at these scenarios to reflect the relevant cognitive and epistemological aspects of the agent's a priori reasoning.

Now we might be told that there are relevant subspaces in Extreme Epistemic Space that have all the structure that I am requesting.[23] For instance, one might observe that there are classes of scenarios in Extreme Epistemic Space that verify A and B whenever they verify $(A \wedge B)$. To make progress on the content and rationality problems, we can simply focus on this class of

[23] Thanks to Jonathan Schaffer for discussion here.

scenarios 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 scenarios in Extreme Epistemic Space that can do the relevant work that we want models of non-ideal epistemic space to do. Rather, we want a method or a procedure for isolating the relevant class of scenarios.

But if one prefers, one can think of subsequent models of non-trivial epistemic space as attempts to formally delineate subspaces in Extreme Epistemic Space that can help us shed light on the content and rationality desiderata. In fact, since nearly all subsequent models of non-ideal epistemic space will inherit the basic properties of Extreme Epistemic Space, there is a relevant sense in which we can think of non-trivial epistemic spaces as subspaces of Extreme Epistemic Space. Yet the conclusion remains that the *structural* features of Extreme Epistemic Space are inadequate for solving the content and rationality problems. Extreme Epistemic Space enables us to model extremely non-ideal agents, but it does not allow us to make sense of the broad class of agents that are not extremely non-ideal.

Though I have used the content and rationality desiderata to illustrate the limitations of Extreme Epistemic Space, we can point to a more general limitation: If we want to use epistemic space to establish familiar modal claims along the lines of "If such-and-such obtains at w, then such-and-such obtains at w", then this space needs to have more structure than Extreme Epistemic Space. For any A, we are trivially guaranteed that there is a scenario w in Extreme Epistemic Space that verifies A. But for any B that stands in some desired relation to A—whether this be a logical, a priori, or conceptual relation—there is nothing in the construction that can guarantee that B will be true at w whenever A is true at w. For any A and B, that is, the construction always allows that A can be true at w while B is false at w. So unless Extreme

Epistemic Space is restricted in some non-arbitrary way, any claim of the form "If such-and-such obtains at w, then such-and-such obtains at w" is prone to counterexamples. In this sense, Extreme Epistemic Space is useful when complete lack of structure is required to model a given phenomenon, but of little use when minimal structure is required. Again, maybe subspaces of Extreme Epistemic Space will contain the relevant structure, but Extreme Epistemic Space itself does not.

Given this, I trust that there is good reason to investigate constructions of non-trivial epistemic spaces that impose substantive constraints on non-ideal scenarios. Extreme Epistemic Space has its roles to play, but if we want a less trivial model of non-ideal epistemic space that can help us satisfy the content and rationality desiderata, we need to go beyond Extreme Epistemic Space—or at least appropriately far inside Extreme Epistemic Space.

2.5 Summary

We have seen that Extreme Epistemic Space is a suitable epistemic space for drawing maximally fine-grained distinctions among epistemic possibilities, and derivatively for modeling extremely non-ideal agents. But just as Ideal Epistemic Space is unsuited for modeling non-ideal agents, so Extreme Epistemic Space is unsuited for modeling the broad class of agents that are not extremely non-ideal. In particular, we need models of non-ideal epistemic space that impose substantive constraints on scenarios to make progress on the content and rationality problems.

In turn we need a notion of deep epistemic possibility that is less trivial than the maximally liberal notion that grounds Extreme Epistemic Space, but less restrictive than the notion that grounds Ideal Epistemic Space. With such a notion of deep epistemic possibility, we can then aim to set up a space of scenarios with a less trivial structure than Extreme Epistemic Space and hopefully make progress on the content and rationality problems.

Accordingly, the main task in the remaining chapters is to investigate constructions of non-ideal epistemic spaces that can avoid the Charybdis of logical omniscience in Ideal Epistemic Space and the Scylla of "anything goes" in Extreme Epistemic Space. In the next chapter, I will lay out the general background structure for models of non-trivial or non-extreme epistemic spaces that can navigate between Ideal Epistemic Space and Extreme Epistemic Space.

2.6 Appendix: A Construction Without the Language and Context Assumptions

For constructions of epistemic spaces in this project, I invoke the language and context assumptions and take the objects of acceptance and rejection to be sentence types in $\mathcal{L}$—that is, sentence types in English*. As a result of this, we can identify scenarios with sets of sentence types in $\mathcal{L}$, and by the simple (Truth) and (Falsity) definitions say that a type A is true (false) at scenario w just in case $A \in w$ ($A \notin w$).

When we give up the language and context assumptions, we take the primary objects of acceptance and rejection to be (possible) sentence *tokens* of arbitrary (existing) natural languages.[24] In contrast to sentence types in natural languages, we can take sentence tokens to have all their broadly context-dependent features fixed on a given occasion of utterance. So here we can say that an agent a accepts (rejects) a sentence token A just in case a is disposed to occurrently accept (reject) A—as above, there are various options for refin-

[24] Though I will here focus on a construction of scenarios that allows us to give up both the language and context assumptions, many of the details apply straightforwardly to a construction of scenarios that only gives up the language assumption. That is, much of the material in this appendix applies to a model—cf. section 2.1.1 above—that constructs scenarios in a single language $\mathcal{L}^+$ and that invokes a translation relation between sentence types in $\mathcal{L}^+$ and sentence types in another given language $\mathcal{L}_i$ in order to evaluate sentence types in $\mathcal{L}_i$ at scenarios constructed in $\mathcal{L}^+$.

ing this definition. So for the purposes of the current construction we should understand quantification over sentences in the relevant (EP) and (Epi-Pos) type principles to involve quantification over (possible) sentence tokens.

We now need a construction of scenarios in a language $\mathcal{L}^+$ that allows us to evaluate arbitrary sentence *tokens* at scenarios. By the reasoning in section 2.1.1 above, there are good reasons not to identify scenarios with either classes of sentence types of natural languages or with classes of possible sentence tokens. Rather, we need a construction of scenarios as sets of sentence *types* in a language $\mathcal{L}^+$ that allows us to evaluate arbitrary sentence tokens of natural languages at these scenarios.

To develop such a construction, we can stipulate a *translation relation* that relates sentence tokens to sentence types in $\mathcal{L}^+$.[25] Generally, the translation relation must have two properties. First, it relates all sentence tokens to corresponding sentence types in $\mathcal{L}^+$. So the translation relation will work across different natural languages. Second, it is sufficiently fine-grained. Intuitively, this means that the translation relation is sensitive to the fine-grained cognitive and epistemological differences that the current notions of epistemic possibility and necessity require. So, for instance, the translation relation must be sufficiently fine-grained to relate tokens of a mathematical truth such as '$1+1=2$' and tokens of a distinct mathematical truth such as '$2+2=4$' to different sentence types in $\mathcal{L}^+$.

So we postulate a language $\mathcal{L}^+$ for which it holds that every possible sentence token is translatable into some sentence type in $\mathcal{L}^+$. To ensure that sentence types in $\mathcal{L}^+$ inherit the epistemic properties of the sentence tokens that they translate, we can say that an agent a accepts (rejects) a sentence type A^+ in $\mathcal{L}^+$ just in case a accepts (rejects) a sentence token that is trans-

[25] When a sentence type in $\mathcal{L}^+$ is a translation of a token A, I will also say that A^+ *translates* A.

lated by A^+ and does not reject (accept) any token that is translated by A^+. It then follows that no agent can simultaneously accept and reject a sentence type A^+ in $\mathcal{L}^+$. We can then say that type A^+ in $\mathcal{L}^+$ is epistemically necessary (impossible) for a just in case a accepts (rejects) A^+. It then follows that no sentence type A^+ in $\mathcal{L}^+$ is both epistemically necessary and epistemically impossible for any agent. So these definitions ensure that if A^+ in $\mathcal{L}^+$ is epistemically necessary (impossible) for a, then no tokens that A^+ translates are epistemically impossible (necessary) for a and at least one token that A^+ translates is epistemically necessary (impossible) for a.

In contrast to the definitions of acceptance and epistemic possibility in section 2.1 above, the current definition of what it means to accept a sentence type in $\mathcal{L}^+$ associates only epistemic possibility with types in $\mathcal{L}^+$ indirectly. On the current construction, the primary objects of acceptance and rejection—and hence of epistemic possibility and necessity—are sentence tokens of arbitrary natural languages. Further, since multiple tokens can be translated into the same type in $\mathcal{L}^+$, the definition above allows that a single type in $\mathcal{L}^+$ can be epistemically necessary for agents that accept and reject tokens of types of different languages. So although tokens are tied to specific agents, the epistemic space that results from constructing scenarios in $\mathcal{L}^+$ need not longer be completely agent-relative.

Given the notion of thoughts from section 2.1.2 above, we might also try to refine the definition of what it means to accept a sentence type in $\mathcal{L}^+$ as follows: An agent a accepts (rejects) a type A^+ in $\mathcal{L}^+$ just in case a accepts (rejects) the content of a thought expressed by a token that is translated by A^+ and does not reject (accept) the content of any thought expressed by a token that is translated by A^+. Insofar as sentence types in $\mathcal{L}^+$ should be specified uniquely by the contents of the associated thoughts of the tokens that they translate, we might also simply say that an agent a accepts (rejects) a type A^+

in $\mathcal{L}^+$ just in case a accepts (rejects) the content of a thought expressed by a token that is translated by A^+. Since no agent can simultaneously accept and reject the content of a thought, it can then never happen that an agent accepts the content of a thought expressed by a token that is translated by A^+ and also rejects the content of a thought expressed by a token that is translated by A^+. We can then also abbreviate 'accept the content of a thought expressed by a token that is translated by A^+' as 'accept the content of a thought expressed by A^+'. Of course, agents do not express thoughts by uttering sentences in $\mathcal{L}^+$, but we can still hold that they can accept a thought the content c of which a speaker of $\mathcal{L}^+$ would express by uttering A^+. As such, we can say that an agent a accepts (rejects) a type A^+ in $\mathcal{L}^+$ just in case a accepts (rejects) the content of a thought expressed by A^+—that is, just in case a accepts (rejects) a thought the content c of which a speaker of $\mathcal{L}^+$ would express by uttering A.

These refinements clearly rely on some notion of 'sameness of content of the thoughts expressed by different tokens'. From what I said in section 2.1 above, I think we have an intuitive grip on what this notion amounts to—below I discuss related matters concerning the translation relation. But if we do not want to rely too much on such a notion, we can still use thought content to characterize intuitively what it means to accept a sentence token and stick to the general definition above: An agent a accepts (rejects) a type A^+ in $\mathcal{L}^+$ just in case a accepts (rejects) a token that is translated by A^+ and does not reject (accept) any token that is translated by A^+. Then we could still give thought content a role to play in a characterization of what it means to accept a sentence type in $\mathcal{L}^+$.

Given this, we can now identify the scenario-making language with the class of sentence types in $\mathcal{L}^+$ and construct scenarios as arbitrary sets of sentence types in $\mathcal{L}^+$. Since the language $\mathcal{L}^+$ is now distinct from the languages in which agents accept and reject sentence tokens, we have to define what it means for a sentence token to be true at a scenario constructed in $\mathcal{L}^+$. The following definition will do the work:

(**Truth**$^+$) A token A is true at scenario w iff the type $A^+ \in w$, where A^+ in $\mathcal{L}^+$ translates A.

(**Falsity**$^+$) A token A is false at scenario w iff the type $A^+ \notin w$, where A^+ in $\mathcal{L}^+$ translates A.

Since each token has a translation in $\mathcal{L}^+$, we can immediately derive (Basic Maximality). And with tiny alterations, we can prove (Epi-Pos) and (Epi-Nec) as we did above.[26]

Since the translation relation is stipulated to relate all sentence tokens to corresponding sentence types in $\mathcal{L}^+$, scenarios constructed in $\mathcal{L}^+$ ensure that we can model agents that accept and reject sentence tokens of all natural languages. Since the translation relation is also stipulated to be sufficiently fine-grained, and since types in $\mathcal{L}^+$ inherit the epistemic properties of the tokens that they translate, scenarios constructed in $\mathcal{L}^+$ also ensure that (Epi-Pos) and (Epi-Nec) remain plausible principles for non-ideal agents. So in its abstract form, the current construction allows us to give up the language and context assumptions.

The current construction leaves open the nature of the translation relation, and as such it leaves open the nature of the language $\mathcal{L}^+$. We could of course take the translation relation as primitive, and let the structural features of the construction do the work for the analyses of epistemic possibility. But clearly, this would not be very informative for understanding the nature of non-ideal epistemic space and the corresponding analyses of epistemic possibility. So

[26] The minor alterations to the proof of (Epi-Pos) and (Epi-Nec) merely consist in replacing all occurrences of '$A \in w$' and '$A \notin w$', where A is a token, with 'A is true at w' and 'A is false at w'. This is enough to incorporate the (Truth$^+$) and (Falsity$^+$) definitions in the background. And as in Extreme Epistemic Space, we can also easily invoke the definition of what it means for two scenarios to be equivalent and derive (Parsimony).

below I discuss an interpretation of the translation relation, and some of the problems that arise when we attempt to explicate such an interpretation.

2.6.1 An Interpretation of the Translation Relation

If we want to give an interpretation of the translation relation, we must specify it in a way that ensures that (Epi-Pos) and (Epi-Nec) remain plausible principles for non-ideal agents. Unsurprisingly, this requires a very fine-grained specification of the translation relation:

First, the translation relation must be sufficiently fine-grained to differentiate between tokens of arbitrary logical and mathematical truths (falsehoods) such as '$1+1=2$' and '$2+2=4$'. For instance, a token A_1 of '$2+2=4$' may remain epistemically possible (impossible) for an extremely non-ideal agent a, even though a token A_2 of '$1+1=2$' is epistemically necessary for a. To ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for such an agent, we then need scenarios in epistemic space that falsify A_1 but verify A_2. By (Truth$^+$), if A_1 and A_2 were translated by the same type A^+ in $\mathcal{L}^+$, then A_1 would be true at w whenever A_2 is true at w. By (Epi-Nec), then A_1 would be epistemically necessary for a whenever A_2 is epistemically necessary for a. So to model an agent for whom A_2 is epistemically necessary while A_1 is epistemically possible (impossible), A_1 and A_2 must be translated by different types A_1^+ and A_2^+ in $\mathcal{L}^+$. So more generally, we can say that tokens of orthographically distinct logical or mathematical truths (falsehoods) within a given natural language typically are translated by different types in $\mathcal{L}^+$, whereas tokens of orthographically identical logical or mathematical truths (falsehoods) typically are translated by a single type in $\mathcal{L}^+$.

Second, the translation relation must be sufficiently fine-grained to differentiate between tokens of trivial truths such as 'Water is water' and 'Hesperus is Hesperus' and tokens of a posteriori necessities such as 'Water is H_2O' and

'Hesperus is Phosphorus'. For instance, a token A_1 of 'Hesperus is Phosphorus' may remain epistemically possible for an astronomically ignorant agent a, even though a token A_2 of 'Hesperus is Hesperus' is epistemically necessary for a. So A_1 and A_2 must be translated by different types A_1^+ and A_2^+ in $\mathcal{L}^+$. If they were not, A_1 would be epistemically necessary for a whenever A_2 is epistemically necessary for a.

Since the current analyses of epistemic possibility apply to these kinds of cases, the following seems clear: The translation relation between sentence tokens and sentence types in $\mathcal{L}^+$ must preserves the fine-grained structure of something akin to *Fregean senses.* In fact, Fregean senses distinguish themselves by having the required level of grain to capture the fine-grained cognitive and epistemological differences that are reflected in the cases above. Roughly, we can think of the Fregean "sense of an expression as mirroring the expression's role in reason and cognition", and we can say that two sentence tokens have the same sense when they play or almost play the same role in reason and cognition.[27]

In some cases, however, there is reason to require that the translation relation is more coarse-grained than Fregean senses:

> Recall that Frege held that the sense of a sentence has an absolute truth-value. This entails that if two utterances of a sentence express the same sense, they must have the same truth-value. But it is clear that certain indexical sentences, such as 'It is now Saturday' can be uttered truly at one time and falsely at another time. So on Frege's picture, these two sentences must have different senses.[28]

[27] Chalmers (2002a): p. 139. In the following, I assume an antecedent, rough grasp of Fregean senses, and trust that the specific cases I discuss below accord well with the (philosophically) ordinary notion of Fregean senses. Also, to be sure, we cannot individuate tokens merely by the orthography of the sentence types that they are tokens of. Since we now aim to deal with various issues concerning context-dependence, tokens of a context-dependent type are often translated by several distinct sentence types in $\mathcal{L}^+$. As such, the translation requires more than mere orthographic individuation of sentence tokens.

[28] Chalmers (2002a): p. 154.

Yet intuitively, though my utterance of 'It is now Saturday' is false today but true if I utter the sentence tomorrow, it is plausible to hold that I in some sense say the same thing or express the same content on both occasions. On both occasions, for instance, I might have no clue about what day it is. To reflect this intuitive notion of same-saying in the translation relation, we can say that typical tokens of many indexical types such as 'It is now Saturday' and 'I am a philosopher' are translated by the same type in $\mathcal{L}^+$. Since tokens of these indexical types have different senses, the translation relation is then more coarse-grained than Fregean senses in these cases.[29]

In other cases, it is possible to require that the translation relation is more fine-grained than Fregean senses. To see this, consider a token A_1 of 'Lawyers are rich' and a token A_2 of 'Attorneys are rich'. Whereas tokens of such intuitively synonymous sentence types plausibly have the same Fregean sense—if anything, A_1 and A_2 seem to play the same role in reasoning and cognition—we might have reason to hold that an extremely non-ideal agent can accept A_1 but nevertheless reject A_2. If so, then the translation relation should relate A_1 and A_2 to different types in $\mathcal{L}^+$, in which case the translation relation should be more fine-grained than Fregean senses.

Yet, there is also an intuition that an agent that accepts a token of 'Lawyers are rich' but simultaneously rejects a token of 'Attorneys are rich' is not really irrational, but rather linguistically ignorant or confused about the meanings of 'lawyer' or 'attorney'. Intuitively, if a speaker really understands and uses 'lawyer' and 'attorney' with full linguistic competence, then she never sincerely dissents from 'Attorneys are rich' when she assents to 'Lawyers are rich'. Partly because my main focus is on non-ideally *rational* agents, but primarily because

[29] Also, if there are tokens of types that involve demonstrative terms such as 'that' and 'there' and that intuitively say the same thing or express the same content, even if they have different truth-values, the translation relation might also be more coarse-grained than Fregean senses in those cases.

of the structure of the translation relation, I will accept this intuition without further ado. This means that I am idealizing away from a certain sort of—depending on intuitions—cognitive or linguistic deficiency that certain agents may suffer from. Of course we must be careful with such an idealization in the current setting. But if we only apply the idealization in cases that involve the standard kinds of intuitively synonymous types such as ‘lawyer’ and ‘attorney’ and ‘vixen’ and ‘female fox’, we should be on safe ground. So granted this idealization, I will say that all tokens of sentence types that only differ in substitution of intuitively synonymous expressions are translated by a single sentence type in $\mathcal{L}^+$. As such, I can hold that the translation relation need never be more fine-grained than Fregean senses. This has the benefit that the translation relation will have a more interesting structure.

A final issue concerns tokens of sentence types in different natural languages that intuitively say or mean the same thing. I want to say that when an English speaking agent a accepts a token E_1 of ‘Two plus two equals four’, and when a German speaking agent b accepts a token G_1 of ‘Zwei plus zwei gleich vier’, they accept the same thing, namely that $2+2=4$. Here the motivation is that we want a construction of scenarios that allows us to define non-ideal epistemic intensions that are useful for modeling the *contents* of the epistemic states of various agents. For that purpose, we want tokens such as E_1 and G_1 to be translated into the same type in $\mathcal{L}^+$. We can again appeal to something akin to Fregean senses to ensure this: Typical tokens of types such as ‘Two plus two equals four’ and ‘Zwei plus zwei gleich vier’ plausibly have the same Fregean sense. Roughly, when used with full competence, such tokens play the same or almost the same roles in reasoning and cognition, or they are composed of terms that play the same or almost the same inferential or conceptual roles in reasoning and cognition. Given this rough characterization, we can then say that when tokens A_1 and A_2 of types in arbitrary natural languages play the same role in reasoning and cognition, then A_1 and A_2 are translated by a single sentence type in $\mathcal{L}^+$. Then E_1 and G_1 above will be translated by the same type in $\mathcal{L}^+$.

This is a very vague characterization of the translation between tokens of types in different natural languages and sentence types in $\mathcal{L}^+$. But I hope the intuitive idea is clear enough: Sentence tokens of types that are obviously synonymous or obviously say the same thing—such as 'Two plus two equals four' and 'Zwei plus zwei gleich vier'—are translated by the same sentence type in $\mathcal{L}^+$. Alternatively, appealing to an antecedent understanding of Fregean senses, whenever sentence tokens obviously have the same Fregean sense, then they are translated by the same sentence type in $\mathcal{L}^+$.

So for the interpretation of the translation relation, the comments above should make it clear that sentence tokens currently must be associated with entities that are akin to Fregean senses. We can put this by saying that the translation relation preserves *quasi-Fregean sense* identity—only *quasi*-Fregean because of the indexical cases above. Since a quasi-Fregean individuation of tokens can do the jobs above, a translation relation that preserves quasi-Fregean sense identity will hence ensure that (Epi-Pos) and (Epi-Nec) remain plausible principles for non-ideal agents.

Before I discuss the issues that arise from the recourse to Fregean senses, I will first fill in the remaining details that are left open by the abstract specification of the construction above. For that purpose, though other options are available, I will employ the notion of thoughts that I introduced in section 2.1.2 above. Thoughts, on the picture above, correspond to token mental states. By assuming a general relation of expression, we can say that each assertive sentence token expresses a thought, where it is a constraint on the relation of expression that the truth-values of the token and the thought always coincide.

We can now stipulate that thoughts are individuated as finely as quasi-

Fregean senses. We can then say that two thoughts t and t' are the *same thought* when t and t' have the same quasi-Fregean sense. Since thoughts correspond to token mental states, this should of course not be understood as numerical identity, but rather, as above, in terms of the roles that these thoughts play in reasoning and cognition. Roughly, if t and t' play the same role in reasoning and cognition, or if t and t' are composed of concepts that play the same inferential or conceptual roles in reasoning, then t and t' are the same (kind of) thought.

We can then let thoughts play the role of the quasi-Fregean entities that we associate with sentence tokens. In particular, we can define two tokens A_1 and A_2 to be *t-equivalent* when A_1 and A_2 express the same thought. Since each token A_i expresses a thought, each A_i will then fall into an equivalence class Σ_{t_i} under the t-equivalence relation. Then each token $A_i \in \Sigma_{t_i}$ expresses the same thought. We can then specify the translation relation as the relation that maps every token $A_i \in \Sigma_{t_i}$ to a distinct sentence type A_i^+ in $\mathcal{L}^+$.

Finally, we can then specify $\mathcal{L}^+$ as the class of sentence types such that each sentence type A_i^+ in $\mathcal{L}^+$ is a translation of each token $A_i \in \Sigma_{t_i}$. Intuitively, we can think of $\mathcal{L}^+$ as an *ideal language*. For in contrast to many types, and particularly in contrast to many context-dependent types in natural (or non-ideal) languages, each (orthographically) distinct sentence type in $\mathcal{L}^+$ is associated with exactly one kind of thought.[30]

The interpretation above ensures that the translation relation can do its jobs. First, the interpretation ensures that the translation relation works for arbitrary tokens of types across natural languages. And second, since thoughts are as fine-grained as quasi-Fregean senses, it ensures that the translation

[30] If we restrict our attention to finite reasoners, "it is plausible that there are only a countable number of relevantly distinct cognitive states." (Chalmers (forthcoming): p. 39.) Following Chalmers, this is particularly plausible if we assume that any finite reasoner can be computationally described. If so, then $\mathcal{L}^+$ contains a countable number of distinct sentence types corresponding to a countable number of distinct thoughts.

relation is sufficiently fine-grained to make (Epi-Pos) and (Epi-Nec) plausible principles for non-ideal agents. So the current interpretation of the abstract construction allows us to use scenarios constructed in $\mathcal{L}^+$ to model agents that accept and reject arbitrary tokens of types in arbitrary natural languages.

But the recourse to Fregean senses primes a serious circularity worry in the current setting. In particular, recall that one of the declared potential applications of non-ideal epistemic space is to motivate a notion of propositional content that although unstructured is as fine-grained as Fregean content. But then, the worries goes, it seems circular to invoke something akin to Fregean senses for the purpose of constructing non-ideal epistemic space. For the current construction of scenarios, a risk of circularity or sense of begging the conclusion is certainly present, and I only have a couple of sketchy remarks to offer in return.

First, impossible worlds are standardly characterized or defined in terms of either sets of sentences or sets of propositions. But it is rarely explained in the literature what the relevant language is or what the relevant notion of propositions is.[31] For an explicit, broadly ersatz construction of impossible worlds that can be used to model Fregean content, I find it hard to see how we can avoid appealing to some notion of content or meaning that plays a role that at least does not conflict with the role that Fregean content plays. Suppose we aim to characterize Fregean content by classes of impossible worlds. For this job, whenever two sentences or thoughts A and B have a different Fregean content, the corresponding construction of impossible worlds must ensure a distinction between A-possibilities and B-possibilities. If we want to give an explicit *construction* of impossible worlds as sets of propositions or interpreted sentences, it seems then as if these propositions or sentences must be individ-

[31] For attempts to develop languages for constructing metaphysically possible worlds, the situation is better; see, for instance, Bricker (1987), Divers (2002), and Lewis (1986).

uated as finely as Fregean content (or something close enough). For instance, if the relevant world-making propositions or sentences were individuated in terms of Russellian propositions, it is very hard to see how there could ever be worlds, so construed, that verify 'Hesperus is Hesperus' but falsify 'Hesperus is Phosphorus'. If this is true, it is plausible that we must assume that the relevant world-making propositions or sentences are individuated or interpreted in terms of entities that are akin to Fregean senses. If so, the risk of circularity not only seems to threaten the construction of scenarios above, but plausibly also many other ersatz constructions of impossible worlds that are engaged in a broadly "Fregean project" and that are explicit about the interpretation of the world-making propositions or sentences.

Second, we might think that Fregean senses are metaphysically mysterious entities, whose job description is nevertheless clear enough. Roughly, as above, Fregean senses are the kinds of entities that allow us to associate different contents with tokens of a priori necessities, and that allows us to associate different contents with tokens of types that only differ in substitution of (many) co-extensional expressions. Given this, we can say that although a construction of non-ideal epistemic space that appeals to Fregean senses cannot justify or ground a Fregean aspect of content, it may nevertheless demonstrate that Fregean senses can be given a precise representation in a formally coherent framework. If successful, this would then allow us to say that Fregean senses need not be more mysterious than other abstract propositional functions.

Third, for the current construction we only need to invoke entities that are *akin* to Fregean senses. For instance, suppose we had (have) a complete metaphysical theory T of the mind—complete in the sense that T would deal with all relevant aspects of the nature of thought (whatever that exactly is). Suppose T argues that when a cognitive system of the relevant kind is in state thus-and-so, then it is in a state that has content c—maybe these states correspond to sentence-like representations in a language of thought. As a consequence of T, suppose it turns out that c plays a role highly reminiscent of Fregean content. And finally, suppose we have good reasons to accept T and to accept a general relation of expression such that every (sincere) utterance expresses a mental state with a content c. Given this, we could then let the mental states of T play the role of Fregean senses and derivatively, it seems, engage in a construction of epistemic space similar to the one above without obvious risk of circularity.[32] That is, *if* we can find independent motivation for a notion of Fregean content in a metaphysical theory of the mind like T, we seem justified in appealing to this notion for a construction of non-ideal epistemic space. Of course, a theory like T would provide immediate justification for a notion of Fregean content, but it could still be of explanatory and theoretical value to embed the notion in a formal framework. So *if* there were (is) a convincing theory of the mind like T, it seems that a construction of non-ideal epistemic space similar to the one above could escape the worst risk of circularity. And at least such a construction would have a well-motivated explanation of why a feeling of circularity arises in the first place.

Clearly, none of the remarks above are able to remove the risk of circularity. And as far as I can tell, this risk remains the biggest—but probably not the only—problem for specifying a relevant interpretation of the translation relation that is compatible with the broadly Fregean project that I am engaged with.

The worries here motivate a more general line of reasoning. We want to use epistemic space to ground a notion of content. To build epistemic space, we need a specification of the translation relation. Insofar as we cannot (innocu-

[32] Of course, the corresponding construction would not be of much interest unless utterance content typically correlates with mental content in some interesting sense—or unless we directly identify the objects of acceptance and rejection with mental states of T rather than sentences.

ously) appeal to content for the specification of the translation relation, the question is then how we can specify an interpretation for it. Although I must leave a detailed discussion of ways to answer this question for future work, let me just very briefly indicate one way of answering the question.

Since we cannot (innocuously) appeal to content for the specification of the translation relation, we might attempt to specify it in broadly behavioristic terms instead. For instance, we might attempt to bootstrap a specification of the translation relation by investigating the kinds of answers that (competent) speakers would give to questions concerning which sentence tokens say or mean the same thing. It is not entirely implausible to imagine that all (or a weighted most of) speakers of English would hold, upon being asked, that tokens of the type A_t 'Attorneys are rich' say or mean the same thing as tokens of the type L_t 'Lawyers are rich'. If so, we could say that the translation relation should relate all tokens of the types A_t and L_t to the same sentence type A^+ in $\mathcal{L}^+$. To ensure that the translation relation could work across different languages, we can imagine asking bilingual or multilingual agents questions concerning which tokens of types of different natural languages say or mean the same thing. It is not entirely implausible to imagine that all bilingual speakers of English and German would hold, upon being asked, that tokens of the English type E_t 'Two plus two equals four' say or mean the same thing as tokens of the German type G_t 'Zwei plus zwei gleich vier'. If so, we could say that the translation relation should relate all tokens of the types E_t and G_t to the same sentence type B^+ in $\mathcal{L}^+$.

By adopting an approach along the lines above, there seems to be hope that we can specify some interesting properties of the translation relation without relying on notions of content or meaning. Yet, several questions and problems are bound to arise. In particular, it is far from obvious how the approach will deal with tokens of types that involve epistemically variant terms such as ‘that’, ‘bank’, and ‘Peter’. For instance, it is unclear which questions we should ask agents to determine when they regard different tokens of types such as ‘Peter is not yellow like that’ as saying or meaning the same thing. We might attempt to overcome such problems by reformulating tokens of epistemically variant types in terms of tokens of epistemically invariant types in the vicinity. But there are plenty of tokens of epistemically variant types for which it is very unclear how we should do this—particularly unclear for those types that involve names and demonstratives.

So when we try to specify an interpretation of the translation relation, very difficult questions and problems quickly arise for translation-based constructions of non-ideal epistemic space. For the main purposes of this project, however, I retain the language and context assumptions. As such, I will in the remaining parts of the thesis always be quantifying over sentence types in $\mathcal{L}$—that is, over sentence types in English*—and always retain the simple construction of scenarios as sets of sentence types in $\mathcal{L}$.

Chapter 3

Non-Trivial Epistemic Spaces

We know that Ideal Epistemic Space and Extreme Epistemic Space are unsuited for capturing facts about epistemic possibility for the broad class of agents that are not extremely non-ideal. In this chapter, I investigate the background structure for models of non-trivial or non-extreme epistemic spaces that can navigate between Ideal Epistemic Space and Extreme Epistemic Space.

To this end, I first motivate a notion of deep epistemic possibility that can be less trivial than the notion that grounds Extreme Epistemic Space, but less restrictive than the notion that grounds Ideal Epistemic Space. Second, I provide a test case interpretation of this notion of deep epistemic possibility in terms of a notion of provability in n steps in some formal system. Third, I use these conceptual tools to lay down the general world involving structure that will motivate subsequent constructions of non-trivial epistemic space. Finally, to avoid confusion, I briefly contrast standard analyses of modal operators with the intended analysis of the provability-in-n-step operator.

By the end of this chapter, we will know the minimal role that non-trivial epistemic space should play to constitute an appropriate framework for modeling the broad class of agents that are not extremely non-ideal.

3.1 Blatant and Subtle Inconsistencies

To capture facts about epistemic possibility for the broad class of agents that are not extremely non-ideal, we need a construction of a non-trivial epistemic space where not “anything goes”. To develop such a space, we need a notion of deep epistemic possibility that is less trivial than the extremely liberal notion that grounds Extreme Epistemic Space, but less restrictive than the ideal notion that grounds Ideal Epistemic Space.

To motivate such a notion of deep epistemic possibility, we can invoke an intuitive distinction between *blatant* and *subtle* inconsistencies. In a letter to Priest, Lewis provides the following piece of introspective evidence:

> 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.[1]

Whereas the sentence ‘$2 + 2 = 5$’ is blatantly inconsistent or impossible on most standards, the falsity of Fermat's Last Theorem is subtly inconsistent or impossible on most standards. On one very natural understanding of what it means for a sentence to be blatantly inconsistent, it means to be a sentence that can easily be rejected by purely a priori reasoning. Intuitively, whereas the thought expressed by a blatantly inconsistent sentence such as ‘$2+2 = 5$’ can be rejected by a cognitively trivial chain of a priori reasoning, this is not the case for thoughts expressed by subtly inconsistent sentences such as ‘There are integers $a; b; c; n > 2$ such that $a^n + b^n = c^n$’.

[1] Lewis (2004): p. 176.

Blatantly inconsistent sentences are the kinds of sentences that any ordinary, reflective reasoner can easily come to reject independently of what empirical information she might happen to have. For instance, if I were to reflect on sentences such as '$0 = 1$', 'My bike is red and blue all over' and 'It rains and it does not rain', I do not have to consult the world to make sure that these sentences are always false. Rather, if I were to reflect on these sentences, I would immediately reject them as false by a cognitively trivial chain of a priori reasoning. Such sentences, we can say, should be and most often are rejected by all ordinary reasoners upon reflection.

In contrast, subtly inconsistent sentences are the kinds of sentences that ordinary, reflective reasoners cannot easily come to reject unless they possess the relevant empirical information. For instance, if I were to suspend all my empirical beliefs and reflect on sentences such as 'One day there will be recursive computers that can prove any mathematical statement that is true' and 'There are integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$', I would not be able to emulate the highly non-trivial chains of a priori reasoning that demonstrate the falsity of these sentences. Ordinary reasoners, we can say, are rationally excused for not rejecting such sentences.

I trust that this picture is intuitive. Insofar as ordinary reasoners are rational at all, blatant inconsistencies are very plausibly the kinds of inconsistencies that we expect them to reject. If a person were to utter '$0 = 1$', 'My bike is red and blue all over' or 'It rains and it does not rain', we would most likely not even deem the person irrational, but instead conclude that he is insincere, confused, or means something different than we do by the relevant utterances. Rather, the cognitive and epistemological aspects of a priori reasoning about blatant inconsistencies are such that any ordinary reasoner, upon reflection, can perform the trivial chain of reasoning that reveals the obvious falsity of these inconsistencies.

We can put these intuitive remarks in terms of epistemic possibility by saying that blatantly inconsistent sentences should always remain epistemically impossible for ordinary, reflective reasoners. More specifically, we can say that blatantly inconsistent sentences always remain epistemically impossible for the class of moderately ideal agents that always engage in the kind of a priori reasoning that ordinary reasoners can easily and often do engage in. By the (EP) analysis of epistemic possibility, we know that if an agent can easily come to reject A by a priori reasoning from what she already accepts and rejects, then A is epistemially impossible for such an agent. Accordingly, since moderately ideal agents have non-trivial cognitive capacities available for instant a priori reasoning, blatantly inconsistent sentences always remain epistemically impossible for these agents, irrespective of what else they might accept or reject, if anything at all.

Since agents are characterized by the cognitive capacities that they have available for instant a priori reasoning, we can now generalize this picture and say that different sentences will count as blatantly inconsistent for different agents. For extremely non-ideal agents we can say that no, or hardly any a priori falsehood counts as blatantly inconsistent. Then every a priori falsehood may remain epistemically possible for such agents. For ideal agents we can say that every a priori falsehood counts as blatantly inconsistent. Then every a priori falsehood always remains epistemically impossible for such agents. For moderately ideal agents we can say that some but not all a priori falsehoods count as blatantly inconsistent. Then some but not all a priori falsehoods always remain epistemically impossible for such agents.

Since deep epistemic possibility is a necessary condition for strict epistemic possibility, we can now use the intuitive picture above to specify a notion of non-ideal deep epistemic possibility that can operate between extreme and ideal deep epistemic possibility. Since I already characterize the distinction between blatant and subtle inconsistencies in terms of easy or obvious a priori reasoning, the following generic definition is natural:

(**G-DEP**) A sentence A is deeply epistemically possible iff A cannot easily be ruled out a priori.

(**G-DEN**) A sentence A is deeply epistemically necessary iff A can easily be established a priori.

By suitable interpretations of the intuitive notion of *easiness*, we can use (G-DEP) and (G-DEN) to isolate a spectrum of notions of deep epistemic possibility, which can range from the maximally liberal to the ideal notion of deep epistemic possibility.

To isolate a maximally liberal notion of deep epistemic possibility, we need an interpretation of *easiness* according to which *no* a priori falsehood can be easily ruled out a priori. The corresponding notion of deep epistemic possibility will be relevant for capturing facts about epistemic possibility for extremely non-ideal agents. To isolate an ideal notion of deep epistemic possibility, we need an interpretation of *easiness* according to which *all* a priori falsehoods can be easily ruled out a priori. The corresponding notion of deep epistemic possibility will be relevant for capturing facts about epistemic possibility for ideal agents. To isolate intermediate notions of deep epistemic possibility, we need an interpretation of *easiness* according to which *some*, but not all a priori falsehoods can be easily ruled out a priori. The corresponding notions of deep epistemic possibility will be relevant for capturing facts about epistemic possibility for the broad class of moderately ideal agents.

To make this picture precise, we need a precise analysis of the notions of *easily rule out a priori* and *easily establish a priori* in (G-DEP) and (G-DEN). In the next section, I offer a test case interpretation that allows us to make these notions precise for the narrow class of a priori truths that are also logical truths.

Corresponding to the spectrum of notions of deep epistemic possibility, the general idea is then to set up a corresponding spectrum of deep epistemic spaces. Roughly, for a maximally liberal notion of deep epistemic possibility, scenarios like those in Extreme Epistemic Space should survive in the corresponding space. Roughly, for an idealized notion of deep epistemic possibility, only scenarios like those in Ideal Epistemic Space should survive in the corresponding space. Roughly, for notions of deep epistemic possibility between these two ends of the spectrum, the corresponding spaces should contain scenarios that are less permissive than those in Extreme Epistemic Space, but less restrictive than those in Ideal Epistemic Space.

The hope is then that we can use scenarios in different spheres in the corresponding stratified epistemic space to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for the whole spectrum of agents. Derivatively, the hope is that we can find a space of scenarios, or a limited spectrum of spaces of scenarios that can help us make progress on the content and rationality problems.

3.2 Provability in n Steps in $\mathcal{S}$

(G-DEP) and (G-DEN) allow us to interpret the notions of *easily rule out* and *easily establish a priori* in various ways. For instance, we might say that a sentence A can easily be ruled out a priori whenever:[2]

(i) it is obvious a priori that $\neg A$; or

(ii) $\neg A$ can be known through such-and-such amount of a priori reasoning; or

(iii) A can easily be disproved by logical reasoning; or

[2] Cf. Chalmers (forthcoming): p. 48.

(iv) $\neg A$ is cognitively insignificant.

In what follows, I will offer an interpretation along the lines of (iii). As a *test case*, that is, I will focus on an interpretation of *easily rule out a priori* as *easily disprovable in some formal system* $\mathcal{S}$ and of *easily establish a priori* as *easily provable in some formal system* $\mathcal{S}$. To be sure, this interpretation does not correspond perfectly to what human agents can easily rule out and establish a priori. Yet, provability in a formal system is a well-understood notion that serves as a good test case for making the intuitive picture above precise. If we cannot construct a suitable non-trivial epistemic space in terms of such simple analyses of (G-DEP) and (G-DEN), it is doubtful whether we can by invoking more complex analyses. So what I say below will generalize to other interpretations of what it means to easily establish and easily rule out a sentence a priori.[3]

In general, I will take proofs to be demonstrations in a formal system $\mathcal{S}$ of the truth of various sentences in languages—including of course English*—that have symbols $\neg$ and $\rightarrow$, which play the same inferential roles as classical negation and material implication. We all know the basic role that a proof in a formal system plays: It gives us a method of reasoning, step by step, according to rules and axioms, to a given conclusion. For simplicity, I will take my canonical system $\mathcal{S}$ to resemble a proof system for standard propositional logic. I will mainly think of $\mathcal{S}$ as a system that contains just axioms and modus ponens, but for purposes of illustration, I will allow myself to use other standard rules of inference such as conjunction-elimination.[4] I will also assume

[3] See chapter 6 for further discussion.

[4] As we know, we can give a full characterization of propositional logic by a system that has just modus ponens and the three axiom schemas:

(A1) $A \rightarrow (B \rightarrow A)$.

(A2) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$.

(A3) $(\neg B \rightarrow \neg A) \rightarrow ((\neg B \rightarrow A) \rightarrow B)$.

that $\mathcal{S}$ is a *sound* system that never proves false sentences.

Given a suitable choice of axioms and rules, we can then define a few key notions. A *proof* in $\mathcal{S}$ is a finite sequence of sentences, each of which either is an axiom of $\mathcal{S}$, or follows from preceding sentences in the sequence via the inference rules in $\mathcal{S}$. A sentence A is a *theorem* of $\mathcal{S}$ or *provable* in $\mathcal{S}$ if and only if there is some proof in $\mathcal{S}$ whose last line is A; by definition, every axiom of $\mathcal{S}$ is hence also a theorem of $\mathcal{S}$. A *disproof* of A in $\mathcal{S}$ is a proof of $\neg A$ in $\mathcal{S}$. A *derivation of A from a set of sentences* Γ in $\mathcal{S}$ is a sequence of sentences such that A is the last sentence in the sequence, and such that each sentence in the sequence is either an axiom of $\mathcal{S}$, a member of Γ, or follows from preceding sentences in the sequence via the inference rules in $\mathcal{S}$. In a derivation in $\mathcal{S}$, sentences that are not theorems may occur. We can also say that every proof of A in $\mathcal{S}$ is a derivation of A in $\mathcal{S}$ from the empty set. A sentence A, or a set of sentences Γ is then *disprovable in $\mathcal{S}$* whenever there is a derivation of A and $\neg A$ in $\mathcal{S}$ from A or Γ.[5] Given these latter definitions, we can then allow $\mathcal{S}$ to operate on assumptions whenever necessary.

Bearing in mind that the *easily provable* and *easily disprovable* interpretations are test cases, we can then interpret *easily provable in $\mathcal{S}$* as *provability in n steps in $\mathcal{S}$*, where n ranges over the natural numbers. A *proof of A in n steps in $\mathcal{S}$* is a proof of A consisting of at most n lines or proof steps in $\mathcal{S}$. A *disproof of A in n steps in $\mathcal{S}$* is a proof of $\neg A$ consisting of at most n steps in $\mathcal{S}$. To *derive A in n steps in $\mathcal{S}$ from a set of sentences* Γ is to have A occurring as the last sentence in some derivation from Γ consisting of at most n steps in $\mathcal{S}$.

For further details, see Hunter (1971): pp. 72-74, and Mendelson (1997): p. 35. Also, as soon as we have $\neg$ and $\rightarrow$, we can express all truth-functions of propositional logic in the standard way.

[5] Since we can express any truth-function of propositional logic in the languages under consideration, we can of course also define or include a symbol $\wedge$ that plays the truth-functional role of classical conjunction. If we do this, we can say that A or Γ is *disprovable in $\mathcal{S}$* whenever there is a derivation of $(A \wedge \neg A)$ in $\mathcal{S}$ from A or Γ. To simplify some of the discussions in chapter 4, however, I will mainly work with the definition in the main text.

A sentence A, or a set of sentences Γ *is disprovable in n steps in $\mathcal{S}$* whenever there is a derivation of A and $\neg A$ from A or Γ consisting of at most n steps in $\mathcal{S}$.

Given these definitions, a step in $\mathcal{S}$ can either consist in an instantiation of an axiom schema in $\mathcal{S}$ or in an application of a rule of inference in $\mathcal{S}$. Let me briefly illustrate. Let $\mathcal{S}$ be an axiomatic proof system characterized by the following two axiom schemas and modus ponens:[6]

(**A1**) $A \rightarrow (B \rightarrow A)$.

(**A2**) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$.

Since $\mathcal{S}$ is an axiomatic proof system, we are not allowed to prove from assumptions. To begin a proof in $\mathcal{S}$, our first step then consists in instantiating an axiom schema. Only afterwards can we start applying modus ponens. To prove $(A \rightarrow A)$, for instance, we may proceed as follows, where the left column tracks proof steps:

(1) $(A \rightarrow ((A \rightarrow A) \rightarrow A)) \rightarrow ((A \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow A))$ (A2)

(2) $A \rightarrow ((A \rightarrow A) \rightarrow A)$ (A1)

(3) $A \rightarrow (A \rightarrow A)$ (A1)

(4) $(A \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow A)$ MP (1,2)

(5) $A \rightarrow A$ MP (3,4)

In the first step, we instantiate axiom schema (A2). In the second and third step, we instantiate (A1). In the fourth and fifth step, we apply modus ponens. Accordingly, $(A \rightarrow A)$ is provable in 5 steps in $\mathcal{S}$, and hence $(A \rightarrow A)$ is a theorem (schema) of $\mathcal{S}$.[7]

[6] For further details and examples, see Bostock (1997): pp. 193-207.

[7] In the case where $\mathcal{S}$ allows for another proof of $(A \rightarrow A)$ that takes more than 5 steps, I will always focus on the 5-step proof. For my purposes, it is always the shortest proof of a given sentence A in $\mathcal{S}$ that matters.

This is the basic idea behind step-based reasoning in $\mathcal{S}$. For my purposes, the exact details of the system $\mathcal{S}$ will not matter much. As long as there is a step-based encoding similar to the one above, the details of $\mathcal{S}$ can vary. However, it is worth making a few more precise remarks about the kinds of systems that I will employ in subsequent chapters.

First, I will restrict my attention to systems that are in some intuitive sense reasonable. For instance, systems in which it takes a trillion steps to employ modus ponens will not be considered. I will not attempt a definition, but simply note that standard proof systems for propositional logic where the procedure behind a n-step proof is familiar will count as reasonable.

Second, I will allow that a single proof in $\mathcal{S}$ can derive both A and $\neg A$. So when a derivation of A and a derivation of $\neg A$ contain the same steps, I will only count these steps once. As such, a single derivation of A and $\neg A$ may contain less steps than a derivation of A plus a derivation of $\neg A$.

Third, for systems that allow derivations from assumptions or premises, I will not count "writing down" premises in a derivation as steps. So to disprove the inconsistent set $\{(\neg A \wedge \neg B), A, B\}$, I do not count writing down first $(\neg A \wedge \neg B)$, then A, and finally B as three initial steps in the relevant derivation.

Finally, I will stipulate that nothing can be proved nor disproved in 0 steps in $\mathcal{S}$. The motivation is here twofold. First, all reasoning takes up cognitive resources. Since I will think of $\mathcal{S}$ as encoding logical reasoning, and think of steps in $\mathcal{S}$ as measuring the cognitive resources that agents have available for logical reasoning—as I will elaborate on below—all logical reasoning thus takes up steps in $\mathcal{S}$. As such, nothing should be provable nor disprovable in 0 steps in $\mathcal{S}$. Second, by holding that nothing can be proved nor disproved in 0 steps in $\mathcal{S}$, scenarios that contain $\{A, \neg A\}$ can survive in some of the constructions of non-trivial epistemic space that I will investigate in chapters 4 and 5. This is desirable if we wish to leave room for agents that may accept both A and

$\neg A$ in models of non-trivial epistemic space. This raises the question of how many steps it should take to disprove sets such as $\{\ldots, A, \neg A, \ldots\}$ in $\mathcal{S}$. In systems that allow derivations from assumptions, I will stipulate that it takes at least 2 steps to disprove such sets. So in order to disprove an inconsistent set $\Delta \cup \{\neg A\}$ by deriving A from Δ using the rules in $\mathcal{S}$, both A and $\neg A$ have to be derived from Δ even when $\neg A$ is already contained in Δ as a premise. Accordingly, I will stipulate that all systems that allow derivations from assumptions have a trivial inference rule that enables us to infer A in 1 step from any set Γ such that $A \in \Gamma$. Given this rule, the inconsistent set $\{(\neg A \wedge \neg B), A, B\}$ from above can then be disproved in 2 steps. First we apply conjunction-elimination on $(\neg A \wedge \neg B)$ to get $\neg A$ in 1 step. Second, we use the trivial rule above and infer A from the set in 1 step. Then A and $\neg A$ can be derived from $\{(\neg A \wedge \neg B), A, B\}$ in 2 steps.

As briefly touched upon, I will think of the system $\mathcal{S}$ as encoding or representing the propositional fragment of a priori logical reasoning. As a general encoding of a priori reasoning, we would need a much broader specification of $\mathcal{S}$. To encode mathematical reasoning, for instance, $\mathcal{S}$ would have to contain at least the axioms of Peano arithmetic—or some rival axiomatization—and hence also first-order logic. To encode what we might call *analytic* or *conceptual* reasoning, $\mathcal{S}$ might have to contain axiom schemas for meaning-postulates; for instance, $\forall(x)(Bx \rightarrow \neg Mx)$, where B and M are predicates for the properties of being a bachelor and being married. To encode *ideal* mathematical and logical reasoning, $\mathcal{S}$ will probably need to contain full second-order logic.[8] Yet, I can make my main points with a simple system that only encodes the propositional fragment of a priori logical reasoning.

[8] In contrast to propositional logic, second-order logic does not yield a complete proof theory and is strongly undecidable. So in contrast to propositional logic, the relationship between logical truth and provability is non-obvious in second-order logic; see Boolos (1975) for further information.

Since various intuitions concerning inferences that are easy in one system but hard in another will not affect my main results, I will not try to compare different encodings of logical reasoning in different systems. For instance, it seems to be a widespread intuition that axiomatic proof systems are harder to use than semantic tableaux systems. But since my main results largely remain unaffected by the choice of proof system—largely in the sense that certain details will differ slightly depending on the choice of proof system—I will continue to encode all relevant logical reasoning in a single system $\mathcal{S}$ similar to a standard propositional system. For what it is worth, the details of some of the results below will be slightly more striking for (humanly) “easy” proof systems like standard semantic tableaux. Further, I will not try to capture the intuition that certain inference rules in a system are harder to employ than others. For instance, it seems to be a widespread intuition that modus tollens is harder to use than modus ponens. In the following, I will primarily focus on modus ponens and rules for conjunction-introduction and elimination. And for what it is worth, these inference rules seem equally easy to employ in reasoning.

More generally, it is worth stressing that the encoding of logical reasoning in a simple formal system like $\mathcal{S}$ only serves as a test case interpretation of the notions of easily establish a priori and easily rule out a priori. Intuitions about what it takes for a piece of a priori reasoning to be ‘easy’ rather than ‘hard’ are likely to differ greatly in many cases, and the interpretation in terms of provability in n steps in $\mathcal{S}$ will definitely not capture all these intuitions. But to evaluate the general prospects of developing a world involving framework for modeling non-ideal a priori reasoning and epistemic possibility, the precise notion of provability in n steps in $\mathcal{S}$ can do the job.

3.3 Epistemic Possibility and Level-n Agents

Given the interpretation of 'easily establish a priori' in terms of provability in n steps in $\mathcal{S}$, let us add the two symbols $\Box_n$ and $\Diamond_n$ to our metalanguage. Let '$\Box_n A$' be read as 'A is provable in n steps in $\mathcal{S}$', and define '$\Diamond_n A$' as $\neg\Box_n\neg A$ and read it as 'A is not disprovable in n steps in $\mathcal{S}$'.[9] We can then state our precise interpretation of the generic notions of deep epistemic possibility and necessity as follows:

(**DEP**) A sentence A is deeply$_n$ epistemically possible iff $\Diamond_n A$.

(**DEN**) A sentence A is deeply$_n$ epistemically necessary iff $\Box_n A$.

That is, A is deeply$_n$ epistemically possible just in case A is not disprovable within n steps in $\mathcal{S}$, whereas A is deeply$_n$ epistemically necessary just in case A is provable within n steps in $\mathcal{S}$. Intuitively, the smaller the value for n in $\Box_n$ is, the easier it is to establish A a priori.

(DEP) and (DEN) can do much of the work that (G-DEP) and (G-DEN) are intended to do. (DEP) and (DEN) immediately give us a spectrum of n notions of deep epistemic possibility. First, by stipulating that no sentence can be disproved in 0 steps in $\mathcal{S}$, we can recover a maximally liberal notion of deep epistemic possibility by letting the value for n in (DEP) be 0. Since no sentence is disprovable in 0 steps in $\mathcal{S}$, no a priori false sentence is deeply$_n$ epistemically impossible when n is 0. Second, by letting the value for n in (DEP) be arbitrarily large, we can *mimic* an ideal notion of deep epistemic possibility.[10] Since all logically false sentences can be disproved in some num-

[9] Strictly, as I will comment upon in a second, $\Box_n$ and $\Diamond_n$ might not count as "real" modal operators, but it does not harm to think loosely of them as such. Also, since I assume that all relevant logical reasoning is encoded by a unique system $\mathcal{S}$, I will not bother about the explicit reference to the system '$\mathcal{S}$' in $\Box_n$—as in the notation '$\Box_n^{\mathcal{S}}$'.

[10] As mentioned, to do more than merely mimic the ideal notion of deep epistemic possibility, we need to enrich the system $\mathcal{S}$ to capture the mathematical and conceptual fragments of a priori reasoning. Though 'All bachelors are unmarried men' is an a priori truth, it will not come out as deeply$_n$ epistemically necessary by (DEN), no matter how large n is.

ber of steps in $\mathcal{S}$, we can ensure that all logically false sentences are deeply$_n$ epistemically impossible when n can go arbitrarily large. Third, by letting the value for n in (DEP) be some sufficiently small (or large) number, we can mimic intermediate notions of deep epistemic possibility. Since some but not all logically false sentences can be disproved in some sufficiently small number of steps in $\mathcal{S}$, we can ensure that some but not all logically false sentences are deeply$_n$ epistemically impossible when n is sufficiently small.

We can think of $\mathcal{S}$ as encoding or representing a priori logical reasoning, and we can think of n as measuring the cognitive capacities that an agent has available for instant or easy a priori reasoning.[11] Intuitively, the larger the value for n is, the larger are the cognitive capacities that an agent has available for instant or easy a priori logical reasoning. We can then say that a *level-n agent a_n* is an agent that can instantly or easily perform up to n steps in $\mathcal{S}$. On this definition, we can then also say that a level-n agent can easily come to accept by a priori reasoning any sentence that is provable within n steps in $\mathcal{S}$, and easily come to reject by a priori reasoning any sentence that is disprovable within n steps in $\mathcal{S}$.[12]

More generally, the definitions above immediately give us the following principles:

($\Diamond_n$**-Level-**n) If A is epistemically possible for a level-n agent a_n, then $\Diamond_n A$.

($\Box_n$**-Level-**n) If $\Box_n A$, then A is epistemically necessary for a level-n agent a_n.

[11] The ideas here bear resemblance to ideas that motivate *step logics*; see for instance Drapkin and Perlis (1986) and Elgot-Drapkin et al. (1991). Notice also that by *instant* or *easy* reasoning, I mean the kind of reasoning that unfolds in relevantly short cognitive episodes for an agent; cf. section 2.1.3, chapter 2.

[12] The ideas here bear resemblance to ideas

For ($\Diamond_n$-Level-n), assume A is epistemically possible for a level-n agent a_n. By the (EP) analysis of epistemic possibility, then a_n cannot easily come to reject A by a priori reasoning. By definition of a level-n agent, a_n can easily come to reject by a priori reasoning any A that is disprovable in n steps in $\mathcal{S}$. So if A is epistemically possible for a_n, then A is not disprovable in n steps in $\mathcal{S}$, and hence $\Diamond_n A$. For ($\Box_n$-Level-n), assume $\Box_n A$. Then A is provable in n steps in $\mathcal{S}$. By definition of a level-n agent, a_n can easily come to accept by a priori reasoning any A that is provable in n steps in $\mathcal{S}$. By the (EN) analysis of epistemic necessity, A is hence epistemically necessary for a_n. If we focus solely on level-n agents that have *no* empirical information, we can also establish ($\Diamond_n$-Level-n) and ($\Box_n$-Level-n) as biconditionals in the obvious way.

Accordingly, the current framework is tailor made to capture the idea that different notions of deep epistemic possibility are relevant for capturing facts about (strict) epistemic possibility for different agents. For instance, consider the class of extremely non-ideal agents. Since extremely non-ideal agents have no cognitive capacities available for instant a priori reasoning, we can characterize these agents as level-0 agents.[13] For any A, if A is epistemically possible for a level-0 agent, then ($\Diamond_n$-Level-n) ensures that $\Diamond_0 A$. By (DEP), then A is deeply$_0$ epistemically possible. Every A is deeply$_0$ epistemically possible, and no A is deeply$_0$ epistemically necessary. Since any A, and in particular any logically false A may remain epistemically possible for extremely non-ideal agents, deep$_0$ epistemic possibility is then the relevant notion of deep epistemic possibility for modeling these agents. Similarly, we can characterize ideal and moderately ideal agents by changing the value for n appropriately and then repeat this reasoning; I will illustrate further in the next chapter.

[13] Or if we want to say that extremely non-ideal agents have few cognitive capacities available for instant a priori reasoning, we can let n be a sufficiently small number.

So I trust that the (DEP) and (DEN) analyses serve as good and precise test cases for capturing the intuitive picture behind models of non-trivial epistemic space. To be sure, (DEP) and (DEN) only allow us to capture facts about which *logically* true or false sentences remain and do not remain epistemically possible for different kinds of agents. But within this narrow a priori domain, we have a well-defined set of conceptual tools that we can use to investigate subsequent constructions of non-trivial epistemic space.

3.4 A Spectrum of Epistemic Spaces

Corresponding to the sequence of provability-in-n-step operators $\Box_0, \Box_1$, $\ldots$, we now need a sequence of deep epistemic spaces $W_0, W_1, \ldots$ that can count as models for the various $\Box_n$ operators. In particular, we need a construction of W_n such that we can establish the following Carnap-style analyses of $\Diamond_n$ and $\Box_n$:

(**C-**$\Diamond_n$) $\Diamond_n A$ iff A is true at some scenario w in W_n.

(**C-**$\Box_n$) $\Box_n A$ iff A is true at all scenarios w in W_n.

If we can find a construction of W_n that allows us to prove (C-$\Diamond_n$) and (C-$\Box_n$), we can derivatively go on to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for the whole spectrum of level-n agents. Since ($\Diamond_n$-Level-n) and ($\Box_n$-Level-n) guarantee a tight connection between deep and strict epistemic possibility, (C-$\Diamond_n$) and (C-$\Box_n$) will immediately ensure that we can use each W_n to capture facts about epistemic possibility for each class of level-n agents; I will go through the details in the next chapter.

Ideally, we want W_n to contain only possible and non-trivially impossible scenarios. Intuitively, we can think of trivially impossible scenarios or blatantly inconsistent scenarios as the kinds of scenarios that any moderately ideal agent can immediately rule out a priori. Non-trivially impossible scenarios or subtly inconsistent scenarios are then the kinds of scenarios that moderately ideal agents cannot rule out in this way. Later, I will define the notion of non-trivially impossible scenarios precisely. But for now we can take the role of non-trivially impossible scenarios to be a placeholder role for a particular class of scenarios in non-trivial epistemic space that we can use to satisfy the content and rationality desiderata. This makes the desideratum that W_n contains only possible and non-trivially impossible scenarios rather loose. Yet, when I evaluate future constructions of non-trivial epistemic space, I hope it will become clear what non-trivially impossible scenarios *cannot* look like to play their roles. At least then we can say which scenarios should *not* survive in W_n if we want to make progress on the content and rationality problems.

To investigate constructions of non-trivial epistemic space, our task is then twofold. First, we need to ensure that we can use the construction to prove (C-$\Diamond_n$) and (C-$\Box_n$). Second, we need to ensure that W_n contains only possible and non-trivially impossible scenarios. If both tasks are met, we can say that the resulting construction can successfully play the role of non-trivial epistemic space.

Before digging into the details of various constructions of non-trivial epistemic space, let me offer a few more comments on the (C-$\Diamond_n$) and (C-$\Box_n$) analyses and the use of the $\Diamond_n$ and $\Box_n$ operators.

3.5 $\Diamond_n$ and $\Box_n$

Though I take (C-$\Diamond_n$) and (C-$\Box_n$) to mirror the core world involving analyses of modal operators in terms of quantification over worlds, a few clarifications are in order to avoid misunderstandings. $\Diamond_n$ and $\Box_n$ are introduced in the metalanguage. This is the reason why I refer to (C-$\Diamond_n$) and (C-$\Box_n$) as 'Carnap-style analyses'. For Carnap, the notion of *L-truth*, which serves as an explication of the notion of logical or analytic truth, is not a term in his object language of state descriptions, but rather in his metalanguage. A sentence A (in the object language) is L-true if and only if A is a member of every state description.[14] I think of the (C-$\Diamond_n$) and (C-$\Box_n$) analyses in a similar way.

Though it might seem as if we are only a short step from a Kripke model for a modal language, I should stress that I am not trying to provide a Kripke model for $\Diamond_n$ and $\Box_n$. To have a proper Kripke model for $\Diamond_n$ and $\Box_n$, we would need to represent $\Diamond_n$ and $\Box_n$ in a modal object language to know what it means for a scenario to satisfy '$\Diamond_n A$' and '$\Box_n A$'. Carnap, in analyzing a modal language, includes the symbol '$N(A)$' in the language with the intended reading 'It is logically necessary that A'.[15] Carnap takes N to be a primitive symbol in the object language that can only be defined in terms of the notion of L-truth in the metalanguage. That is, '$N(A)$' is true if and only if 'A' is L-true.[16] Then a sentence '$N(A)$' is true at a state description S if and only if 'A' is true at all state descriptions S'.

Suppose we tried in a similar way to represent $\Box_n$ as a primitive symbol '$\blacksquare_n A$' in the agent (or object) language, where '$\blacklozenge_n A$' is defined as $\neg\blacksquare_n\neg A$. We would then have:

(C-$\blacklozenge_n$) $\blacklozenge_n A$ is true at a scenario w iff A is true at some scenario w' in W_n.

(C-$\blacksquare_n$) $\blacksquare_n A$ is true at a scenario w iff A is true at all scenarios w' in W_n.

(C-$\blacklozenge_n$) and (C-$\blacksquare_n$) would make $\blacklozenge_n$ and $\blacksquare_n$ satisfy the modal axioms (S4) and

[14] A *state description* is a class of sentences in the object language that contains for every atomic sentence A (in the object language) either A or $\neg A$, but not both; cf. Carnap (1947): pp. 8-11.

[15] See chapter 5 in Carnap (1947) for all the details.

[16] Carnap (1947): p. 174, convention 39-1.

(S5).[17] For Carnap's notion of logical necessity, this is not a problem. But insofar as we can make sense of iterated principles for $\Diamond_n$ and $\Box_n$ without specifying what it means for $\mathcal{S}$ to represent facts about n-provability, it seems clear that neither (S4) nor (S5) should hold. Intuitively, the fact that A is provable in n steps in $\mathcal{S}$ does not entail that this fact too is provable in n steps in $\mathcal{S}$. For (S5), the fact that A is not disprovable in n steps in $\mathcal{S}$ does not entail that this is provably so in n steps in $\mathcal{S}$. Further, since we think of $\mathcal{S}$ as encoding bounded a priori logical reasoning, we should also intuitively expect iteration to fail for $\Diamond_n$ and $\Box_n$ because a priori reasoning about a priori reasoning itself takes up cognitive resources. If so, then '$\blacksquare_n$' in (C-$\blacksquare_n$) does not translate as provability in n steps in $\mathcal{S}$ in our metalanguage.

One might try to block these iterations by amending (C-$\blacklozenge_n$) and (C-$\blacksquare_n$) with an accessibility relation $\mathcal{R}$. To invalidate both (S4) and (S5) for $\blacklozenge_n$ and $\blacksquare_n$, one might stipulate that $\mathcal{R}$ is to be a non-transitive relation on scenarios in W_n. But then we face another problem. If not all $w \in W_n$ are accessible from each other, then '$\blacksquare_n A$' might be true at some w even though A is not provable in n steps. This can happen if all scenarios w' that verify $\neg A$ are inaccessible from w. If A is a contingent sentence, then again '$\blacksquare_n$' would fail to translate as provability in n steps in $\mathcal{S}$ in our metalanguage. There might be ways to define and motivate an accessibility relation on scenarios in W_n that can avoid these problems, but the properties of the relation would have to be highly non-standard from the outset.

Instead, one might attempt to go the way of provability logic and introduce $\Box_n$ in a modal object language, which has a model theory that can ensure that '$\Box_n A$' only holds when A is provable in n steps in $\mathcal{S}$. On such an approach, $\Box_n$

[17] This is easily seen. For (S5), let '$w \vDash A$' abbreviate 'A is true at w', and assume $w \vDash \blacklozenge_n A$. By (C-$\blacklozenge_n$), there is then a w' in W_n such that $w' \vDash A$. Since all scenarios in W_n are accessible from each other, no matter which w' in W_n we pick, it follows from (C-$\blacklozenge_n$) that $w' \vDash \blacklozenge_n A$. So for all w' in W_n, $w' \vDash \blacklozenge_n A$. Then, by (C-$\blacksquare_n$), $w \vDash \blacksquare_n \blacklozenge_n A$. So (S5) is valid: $\blacklozenge_n A \rightarrow \blacksquare_n \blacklozenge_n A$. A similar line of reasoning applies to (S4): $\blacksquare_n A \rightarrow \blacksquare_n \blacksquare_n A$.

would represent the formalized provability-in-n-step predicate of $\mathcal{S}$, and presumably correspond to a rather complex arithmetical formula. An interesting open question is then which modal logic should govern $\Box_n$, and how exactly $\mathcal{S}$ can represent $\Box_n$.[18] If we can make sense of the modal logic for $\Box_n$, we could then attempt to specify a Kripke model for the logic. Just as there is a Kripke model for the standard modal logic $\mathcal{GL}$ of arithmetical provability, so there is probably some kind of Kripke model for the modal logic of $\Box_n$.[19] The final open question is then whether the indices in such a Kripke model would enable us to shed light on broad issues concerning hyperintensionality in philosophy of language and mind.

Both proof-theoretical and model-theoretical properties of $\Diamond_n$ and $\Box_n$ are interesting in their own right, and much more could be said. But for purposes of the current project, the basic formulations (C-$\Diamond_n$) and (C-$\Box_n$) will do. I will continue to refer to $\Diamond_n$ and $\Box_n$ as 'operators', even if "real" modal operators need a proper model theory. Since the $\Diamond_n$ and $\Box_n$ notation primarily plays an expository role, this should be a safe path to follow in investigating constructions of non-trivial epistemic space.

3.6 Summary

To find a successful construction of non-trivial epistemic space, our task for the remaining chapters is clear. We need to find a construction of a stratified

[18] Though there are tricky question concerning how $\mathcal{S}$ can represent $\Box_n$, we will probably get the Löb axiom immediately: $\Box_n(\Box_n A \rightarrow A) \rightarrow \Box_n A$. If so, then either the rule of necessitation or the (T) axiom would have to go on pain of Gödel's theorems. If we give up the rule of necessitation, the resulting modal logic will be non-normal. If we give up (T), we cannot ensure that $\Box_n A$ only holds for true A. So we are most likely left with a non-normal modal logic for the formalized provability-in-n-step-in-$\mathcal{S}$ predicate.

[19] See Boolos (1979) and (1993) for detailed constructions of Kripke models for arithmetical provability. It is a noticeable feature of the standard Kripke models for arithmetical provability that the intuitive meanings of the indices and the transitive and conversely well-founded accessibility relation are left unexplained. Though this sits rather uneasily with the constructive approach to epistemic space that I pursue in this project, maybe one would have to leave the indices and the accessibility relation in a Kripke model for $\Box_n$ unexplained as well.

epistemic space that will satisfy (C-$\Diamond_n$) and (C-$\Box_n$), and that derivatively will ensure that W_n contains only possible and non-trivially impossible scenarios. If we can find such a construction, we will have a non-trivial, yet non-ideal epistemic space whose main job is to capture facts about epistemic possibility for the broad class of moderately ideal agents. If successful, we would be a step closer to finding a construction of a non-ideal epistemic space that can avoid the Scylla of “anything goes” and the Charybdis of logical omniscience.

In the following, I will argue that successful constructions of non-trivial epistemic spaces are hard, if not impossible to find. In particular, whereas it is rather simple to construct W_n in a way that allows us to prove (C-$\Diamond_n$) and (C-$\Box_n$), we will see that the task of ensuring that W_n contains only possible and non-trivially impossible scenarios remains formidable. The problems with the models that I will investigate in the next chapter motivate a vicious dilemma, which threatens to undermine the intuitive picture behind non-trivial epistemic space that I have set up in this chapter. In chapter 5, I will prove that the dilemma holds and discuss the options that remain open for constructions of non-trivial epistemic space in light of this dilemma.

Chapter 4

Constructions of Non-Trivial Epistemic Space

In this chapter, I investigate three intended constructions of non-trivial epistemic space. Though we can use the first model, the Single Disprovability Model, to ensure the Carnap-style analyses of $\Diamond_n$ and $\Box_n$, I argue that the resulting scenarios remain extremely unconstrained and cannot play the role that we want non-trivially impossible scenarios to play. Though we can use the second model, the Joint Disprovability Model, to eliminate the trivially impossible scenarios that survive in the Single Disprovability Model, I argue that all impossible scenarios are eliminated from the resulting epistemic space. Derivatively, we cannot use the construction to model the broad class of moderately ideal agents. As a version of the Joint Disprovability Model, I finally investigate Jago's recent construction of a non-ideal epistemic space in *Logical Information and Epistemic Space*, and argue that agents characterized by this model turn out to be logically omniscient.[1]

By the end of this chapter, we will have seen how models of non-ideal epistemic space threaten to yield either too many trivially impossible scenarios, where almost any set of blatantly inconsistent sentences can be true, or too few impossible scenarios, where any logical falsehood can be true.

[1] Jago (2009a).

4.1 The Single Disprovability Model

To develop a sequence of epistemic spaces that can count as models for the sequence of $\Box_n$ operators, the main strategy behind the *Single Disprovability Model* is to define each epistemic space directly in terms of the notion of provability in n steps in system $\mathcal{S}$. For a given scenario w to be a member of a given epistemic space W_n, the idea is, w cannot verify any sentence that is disprovable within n steps in $\mathcal{S}$. Intuitively, the more steps in $\mathcal{S}$ that it takes to disprove any sentence that is true at the scenario, the more subtly inconsistent the scenario will be.

To construct the Single Disprovability Model, I will use the basic material from the construction of Extreme Epistemic Space. First, a scenario w is identified with an arbitrary set of sentences in $\mathcal{L}$—that is, with an arbitrary set of possible sentence types in English*. Second, a sentence A is true at a scenario w if and only if $A \in w$, and A is false at w if and only if $A \notin w$. Third, two scenarios w and w' are equivalent if and only if for all A, $A \in w$ if and only if $A \in w'$. Additionally, I require that scenarios are *minimally closed* in the following sense:

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

Given this, we can now establish that scenarios obey the following principle of maximality:

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

Since $A \in w$ if and only if $\neg A \notin w$, for each sentence A and scenario w, (Maximality) follows immediately from (Min-Clo), (Truth) and (Falsity). When a scenario w—or a set of sentences more generally—obeys (Maximality), I will also say that w is *maximal.*[2] As in Extreme Epistemic Space, (Parsimony) also follows immediately. Let W_S be the class of scenarios that satisfy these basic principles.

We need (Min-Clo) for the spherical construction of the Single Disprovability Model, and we need (Maximality) to establish the Carnap-style analysis (C-$\Diamond_n$). Since (Min-Clo) effectively equates 'falsity of A at w' with 'truth of $\neg A$ at w' in the classical sense, (Maximality) captures the idea behind (Basic Maximality) that says that for all A and w, either A is true at w or A is false at w. (Min-Clo) also entails that scenarios in W_S are *minimally consistent* in the sense that $\{A, \neg A\} \nsubseteq w$ for any w. As such, we know that all scenarios in W_S verify either A or $\neg A$ but never both.

This of course means that we cannot use scenarios in the Single Disprovability Model to model agents for whom both A and $\neg A$ are epistemically necessary or impossible. But since I am now mainly interested in moderately ideal agents, this is desirable: Moderately ideal agents never accept nor reject both A and $\neg A$.[3] We should notice though that (Min-Clo) does *not* entail that all scenarios falsify $(A \wedge \neg A)$. It merely entails that if w verifies $(A \wedge \neg A)$, then w also falsifies $\neg(A \wedge \neg A)$. In this sense, (Min-Clo) does not prevent us from modeling extremely non-ideal agents for whom arbitrary *explicit* contradictions of the form $(A \wedge \neg A)$ may be epistemically possible or necessary.

We can now construct the sequence of epistemic spaces $\big(W_0, W_1, \dots\big)$ that we need to prove (C-$\Diamond_n$) and (C-$\Box_n$). To this end, let PR_n be the set of sentences that are provable in up to n steps in $\mathcal{S}$. We can then define the sequence of epistemic spaces as follows:

$$W_0 = \{w \in W_S | PR_0 \subset w\}.$$

[2] So when a set of sentences Γ is maximal, this means effectively that for all sentences A, either $A \in \Gamma$ or $\neg A \in \Gamma$.

[3] Cf. section 2.1.3, chapter 2.

$$W_1 = \{w \in W_S | PR_1 \subset w\}.$$

$$W_2 = \{w \in W_S | PR_2 \subset w\}.$$

$$\vdots$$

$$W_n = \{w \in W_S | PR_n \subset w\}.$$

$$\vdots$$

$$W_\infty = \{w \in W_S | PR_\infty \subset w\}.$$

We can stipulate that no sentence A can be proved in 0 steps in $\mathcal{S}$. Then PR_0 is empty, and hence any scenario w in W_S is a member of W_0. So W_S is identical to the class of scenarios W_0. As we increase the value for n, increasingly more non-ideal scenarios will be excluded from W_n according to the recipe: For any $A \in PR_n$, for $n > 0$, and any $w \in W_0$, if $\neg A \in w$, then $w \notin W_n$.[4] W_n then corresponds to the class of scenarios in W_0 whose only constraint is that they verify all sentences that are provable within n steps in $\mathcal{S}$. W_∞ corresponds to the class of scenarios in W_0 that verify *all* provable sentences in $\mathcal{S}$.[5]

The various spheres of scenarios that we can isolate in the Single Disprovability Model are hence related by the subclass relation:

$$W_0 \supseteq W_1 \supseteq W_2 \ldots \supseteq W_n \ldots \supseteq W_\infty.$$

Intuitively, as we move inwards through this stratified epistemic space, scenarios will verify more and more provable sentences, or equivalently, falsify more and more disprovable sentences.

[4] To illustrate, let A be in PR_n, for $n > 0$, and assume $\neg A \in w$. Assume (per impossibile) that $w \in W_n$. By construction, w then contains all $A \in PR_n$. So $A \in w$. But by assumption, $\neg A \in w$, and hence $A \notin w$ by (Min-Clo). So $w \notin W_n$.

[5] Notice that the use of the symbol '∞' should not be taken to imply that we can construct proofs in $\mathcal{S}$ that involve infinitely many steps. Since $\mathcal{S}$ mirrors a standard system for propositional logic, all proofs in $\mathcal{S}$ are of finite length. Rather, I use the symbol '∞' to indicate that we can construct proofs in $\mathcal{S}$ that involve arbitrarily large finite numbers of steps. Since each logically true sentence can be proved in some number of steps in $\mathcal{S}$, PR_∞ then corresponds to the class of sentences that are provable in $\mathcal{S}$.

Given this construction of the Single Disprovability Model, we can now establish (C-$\Diamond_n$) and (C-$\Box_n$):

> **(C-**$\Diamond_n$) $\Diamond_n A$ iff A is true at some scenario w in W_n.
>
> **(C-**$\Box_n$) $\Box_n A$ iff A is true at all scenarios w in W_n.

The proof of (C-$\Diamond_n$) and (C-$\Box_n$) is easy:

> For (C-$\Diamond_n$) left to right, assume $\Diamond_n A$. Since $\Diamond_n A$ is defined as $\neg\Box_n\neg A$, then $\neg A$ is not provable in n steps in $\mathcal{S}$. So $\neg A \notin PR_n$. By construction of W_n, there is then a $w \in W_n$ such that $\neg A \notin w$. By (Maximality), there is hence a $w \in W_n$ such that $A \in w$. For (C-$\Diamond_n$) right to left, assume $A \in w$ for some scenario $w \in W_n$. Then it is not the case that for all $w \in W_n, A \notin w$. By (Maximality), then it is not the case that for all $w \in W_n, \neg A \in w$. By construction of W_n, then $\neg A \notin PR_n$. Then $\neg A$ is not provable in n steps in $\mathcal{S}$. So $\neg\Box_n\neg A$, and hence by definition $\Diamond_n A$. So (C-$\Diamond_n$) holds.
>
> For (C-$\Box_n$) left to right, assume $\Box_n A$. Then A is provable in n steps in $\mathcal{S}$. So $A \in PR_n$. By construction of W_n, then $A \in w$ for all $w \in W_n$. For (C-$\Box_n$) right to left, assume $A \in w$ for all $w \in W_n$. For reductio, assume $A \notin PR_n$. Then, by construction of W_n, there is a $w \in W_n$ such that $A \notin w$. By assumption, this cannot happen, so $A \in PR_n$. Then A is provable in n steps in $\mathcal{S}$, and hence $\Box_n A$. So (C-$\Box_n$) holds. □

Given (C-$\Diamond_n$) and (C-$\Box_n$), we can now use each deep epistemic space W_n in W_0 to make (Epi-Pos) and (Epi-Nec) plausible principles for the whole spectrum of agents:

(**Epi-Pos**) A is epistemically possible for an agent a iff there is a scenario w in W such that w is epistemically possible for a and such that A is true at w.

(**Epi-Nec**) A is epistemically necessary for an agent a iff for each scenario w in W such that w is epistemically possible for a, A is true at w.

On the current picture, agents are characterized by the cognitive capacities that they have available for instant or easy a priori reasoning. As a test case, I focus on the logical fragment of a priori reasoning and characterize each agent as belonging to a class of level-n agents. A level-n agent is an agent that can instantly perform up to n steps in $\mathcal{S}$. So to make (Epi-Pos) and (Epi-Nec) plausible for the whole spectrum of agents, we need to make (Epi-Pos) and (Epi-Nec) plausible for each class of level-n agents.

This requires that the relevant class W of deeply epistemically possible scenarios in (Epi-Pos) and (Epi-Nec) satisfy the following:

(**D1***) If A is epistemically possible for any level-n agent, then there is some scenario in W that verifies A.

(**D2***) If A is true at each scenario in W, then A is also epistemically necessary for each level-n agent.

If W satisfies (D1*), we are guaranteed that W contains enough scenarios to model what is epistemically possible for a given level-n agent. If W satisfies (D2*), we are guaranteed that only sentences that are already epistemically necessary for each level-n agent are true throughout the class of scenarios in W. For (D2*), that is, we are assured that the relevant class of scenarios in W only verify sentences that any level-n agent could easily come to accept by purely a priori reasoning.

By the definition of a level-n agent, we have from the previous chapter:

(**$\Diamond_n$-Level-n**) If A is epistemically possible for a level-n agent a_n, then $\Diamond_n A$.

(**$\Box_n$-Level-n**) If $\Box_n A$, then A is epistemically necessary for a level-n agent a_n.

In conjunction with (C-$\Diamond_n$) and (C-$\Box_n$), we can immediately use ($\Diamond_n$-Level-n) and ($\Box_n$-Level-n) to secure that W_n satisfies (D1*) and (D2*) as:

(**D1**) If A is epistemically possible for any level-n agent, then there is some scenario in W_n that verifies A.

(**D2**) If A is true at each scenario in W_n, then A is also epistemically necessary for each level-n agent.

As a special case, we have (D1) and (D2) as biconditionals for the class of level-n agents that have *no* empirical information. For these agents, we can say that deep and strict epistemic possibility and necessity coincide.

So to make (Epi-Pos) and (Epi-Nec) plausible for each class of level-n agents, we know that $W = W_n$. To capture this explicitly, we can stratify the (Epi-Pos) and (Epi-Nec) principles in the obvious way as follows:

(**Epi-Pos$_n$**) A is epistemically possible for a level-n agent a_n iff there is a scenario w in W_n such that w is epistemically possible for a_n and such that A is true at w.

(**Epi-Nec$_n$**) A is epistemically necessary for a level-n agent a_n iff for each scenario w in W_n such that w is epistemically possible for a_n, A is true at w.[6]

[6] Though the proof of (Epi-Pos) and (Epi-Nec) from chapter 2 still applies to (Epi-Pos$_n$) and (Epi-Nec$_n$), I will quickly prove (Epi-Pos$_n$) to show how the proof of (Epi-Pos) can be simplified in models of epistemic space that satisfy (Maximality); the proof of (Epi-Nec) cannot be simplified. Since we can define epistemic possibility in terms of epistemic necessity, we can simplify the definition of (Epi-Pos-w) as such: A scenario w in W_n is epistemically

I will briefly illustrate how this picture works.

First, for extremely non-ideal agents, any logically false sentence may be epistemically possible. Since such agents have no cognitive capacities available for instant a priori logical reasoning, they are characterized as level-0 agents. By construction, any logically false A is true at some scenario w in W_0. So to make (Epi-Pos) and (Epi-Nec) plausible principles for extremely non-ideal agents, we use (Epi-Pos$_0$) and (Epi-Nec$_0$) and identify W with W_0.

Second, for moderately ideal agents, some but not all logically false sentences may be epistemically possible. Since such agents have some cognitive capacities available for instant a priori logical reasoning, they are characterized as level-m agents, for some sufficiently small (or large) m. By construction, some but not all logically false A are true at scenarios in W_m. So to make (Epi-Pos) and (Epi-Nec) plausible principles for moderately ideal agents, we use (Epi-Pos$_m$) and (Epi-Nec$_m$) and identify W with W_m.

Third, for ideal agents, no logically false sentence may be epistemically possible. Since such agents have unbounded cognitive capacities available for instant a priori logical reasoning, they are characterized as level-∞ agents. By construction, all logically false A are false at each scenario in W_∞. So to make (Epi-Pos) and (Epi-Nec) plausible principles for ideal agents, we use (Epi-Pos$_\infty$) and (Epi-Nec$_\infty$) and identify W with W_∞.

possible for a level-n agent a_n if and only if for all A, if A is epistemically necessary for a_n, then $A \in w$. The proof of (Epi-Pos$_n$) is then:

> For (Epi-Pos$_n$) left to right, assume A is epistemically possible for a_n. Then $\neg A$ is not epistemically necessary for a_n. By (Epi-Pos-w), then it is not the case that $\neg A \in w$ for all $w \in W_n$ that are epistemically possible for a_n. There is then a $w \in W_n$ such that $\neg A \notin w$ and such that w is epistemically possible for a_n. By (Maximality), there is hence a $w \in W_n$ such that $A \in w$ and such that w is epistemically possible for a_n. For (Epi-Pos$_n$) right to left, assume $A \in w$ for some $w \in W_n$ that is epistemically possible for a_n. Then it is not the case that $A \notin w$, and hence by (Maximality) not the case that $\neg A \in w$ for all $w \in W_n$ that are epistemically possible for a_n. Suppose, for reductio, that $\neg A$ is epistemically necessary for a_n. By (Epi-Pos-w), then $\neg A \in w$ for all w that are epistemically possible for a_n. Contradiction. So $\neg A$ is not epistemically necessary for a_n. So A is epistemically possible for a_n. So (Epi-Pos$_n$) holds. □

Accordingly, since (C-$\Diamond_n$) and (C-$\Box_n$) hold in the Single Disprovability Model, we can use the model to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for the whole spectrum of agents. So the Single Disprovability Model meets the first challenge that a successful construction of non-trivial epistemic space needs to meet. For the second challenge, we can then ask whether W_n, for some n, contains only possible and non-trivially impossible scenarios. I will now argue that the Single Disprovability Model fails to meet this second challenge, and hence that the model fails to play the role of non-trivial epistemic space.

4.1.1 Problems in the Single Disprovability Model

Consider a scenario $w \in W_n$. By construction, w verifies all sentences that are provable within n steps in $\mathcal{S}$. Since w is also minimally closed, w can then never verify $\neg A$ (A) when A ($\neg A$) is provable in n steps in $\mathcal{S}$. But aside from this, there is nothing in the Single Disprovability Model that prevents scenarios in W_n from being arbitrarily *jointly* inconsistent. In particular, for any contingent sentences A and B and any epistemic space W_n, W_n can contain scenarios like w and w':

$$\begin{aligned} w = \;& PR_n \cup \{(A \wedge B), \neg A, \neg B, \ldots\}. \\ w' = \;& PR_n \cup \{(A \rightarrow B), A, \neg B, \ldots\}. \end{aligned}$$

For arbitrarily large n, there are scenarios like w in W_n that verify 'It rains and the streets are wet', but also 'It does not rain' and 'The streets are not wet'. And for arbitrarily large n, there are scenarios like w' in W_n that verify 'If it rains, the streets are wet', but also 'It Rains' and 'The streets are not wet'.[7]

[7] For purposes of illustration, I assume here—as elsewhere—that 'It Rains' and 'The streets are wet' are sentence types in $\mathcal{L}$ or English*.

Since we only evaluate whether single sentences are provable or not, the presence of such scenarios in W_n is unavoidable. Intuitively, $\mathcal{S}$ scans each $A \in w$ to determine whether A is provable or disprovable in some number of steps. Yet, since $\mathcal{S}$ is a sound system, there are many contingent sentences in any w that $\mathcal{S}$ neither proves nor disproves, and this no matter how jointly inconsistent these sentences in w might otherwise be. As a result, scenarios in the Single Disprovability Model inherit much of the trivial structure that scenarios in Extreme Epistemic Space have: Except for containing each sentence that is provable within n steps in $\mathcal{S}$, scenarios in W_n can be arbitrarily jointly inconsistent. With respect to the class of contingent sentences, the Single Disprovability Model effectively has no further structure than Extreme Epistemic Space. This is not enough to secure that W_n contains only possible and non-trivially impossible scenarios. To see this, we consider the content and rationality desiderata.

For the content desideratum, we want to use scenarios in W_n to give a world involving account of a non-trivial notion of hyperintensional content. In particular, we want to use non-ideal epistemic intensions to represent the contents of the epistemic states of agents that are *not* extremely non-ideal. To investigate whether the kinds of non-ideal epistemic intensions that we can define in the Single Disprovability Model can play this role, let the *epistemic n-intension* of a sentence be a function from scenarios in W_n to a truth-value. By the construction of W_n, we can immediately ensure that the epistemic n-intension of any n-provable A is necessary. So in contrast to Extreme Epistemic Space, we can isolate certain non-trivial epistemic intensions in the Single Disprovability Model.

But when we look beyond the class of n-provable sentences, the epistemic n-intensions remain completely unconstrained. Consider the simple inference from $(A \wedge B)$ to A, where A and B are contingent sentences. To reflect such basic inferential relations among sentences and thoughts in the corresponding non-ideal epistemic intensions, we cannot use the Single Disprovability Model: Though the n-intension of $(A \wedge B)$ is true at a scenario $w \in W_n$, there is no guarantee that the n-intension of A is also true at w. For all the construction says, it may well be that the n-intension of A is false at w while the n-intension of $(A \wedge B)$ is true at w. But to reflect the inference from $(A \wedge B)$ to A, the n-intension of A should be true whenever the n-intension of $(A \wedge B)$ is true.

Or consider the basic inferential relations between $(A \leftrightarrow B)$ and A and B, where A and B are contingent sentences. Assume $(A \leftrightarrow B)$ is provable within n steps in $\mathcal{S}$. Then $(A \leftrightarrow B)$ is true at each $w \in W_n$. Yet, since A and B are contingent, there are plenty of scenarios in W_n at which the truth-values of A and B can come arbitrarily apart. Derivatively, though the n-intension of $(A \leftrightarrow B)$ is true at all $w \in W_n$, the values of the n-intensions of A and B can come arbitrarily apart at scenarios in W_n. But to reflect the inferential relations between such sentences, the n-intensions of A and B should take the same value when the n-intension of $(A \leftrightarrow B)$ is necessary.

So to represent such cognitively trivial or computationally feasible inferences between thoughts and sentences, we cannot use epistemic n-intensions. So epistemic n-intensions cannot play the role that we want non-ideal epistemic intensions to play to satisfy the content desideratum. To make progress on the content problem, scenarios need to obey further constraints than scenarios in the Single Disprovability Model.

For the rationality desideratum, we want to use scenarios in W_n to give a world involving analysis of a non-trivial notion of epistemic possibility that captures which sentences should remain epistemically possible for minimally rational agents. In contrast to Extreme Epistemic Space, we might be able to use scenarios in the Single Disprovability Model to capture *some* aspects of a notion of minimal rationality. To illustrate, let $\mathcal{X}$ be the set of sentences that can be disproved within, say, 15 steps in $\mathcal{S}$. Suppose we have intuitive reason to hold that each sentence A in $\mathcal{X}$ counts as obviously false and that each moderately ideal agent rationally should reject each $A \in \mathcal{X}$. To give a world involving analysis of a notion of epistemic possibility that captures this aspect, we can then use W_{15}. Since each $A \in \mathcal{X}$ is false at each $w \in W_{15}$, we then know that each $A \in \mathcal{X}$ remains epistemically impossible for any agent modeled by the space W_{15}. Derivatively, we can say that any A that is false at each $w \in W_{15}$ should remain epistemically impossible for any moderately ideal agent, whereas any A that is true at some $w \in W_{15}$ may remain epistemically possible for such agents.

But again, when we look beyond the class of n-provable sentences, the Single Disprovability Model is of little use. In particular, if we want to capture facts about which sentences a minimally rational agent should accept, *given* that she accepts thus and so, the Single Disprovability Model lacks the required structure. For instance, since a moderately ideal agent can easily infer A from $(A \wedge B)$, for some contingent A and B, we want to say that she rationally should accept A when she accepts $(A \wedge B)$. We capture this normative component in the (EN) analysis by saying that if $(A \wedge B)$ is epistemically necessary for a moderately ideal agent, then A is also epistemically necessary for this agent. To analyze such situations, we need to ensure that if $(A \wedge B)$ is true at each $w \in W_n$, for some n, then A is also true at w. The Single Disprovability Model cannot do this job: For *any* epistemic space W_n, though $(A \wedge B)$ is true at each $w \in W_n$ that remains epistemically possible for a given moderately ideal agent, there is no guarantee that A is also true at w. In fact, for all the construction says, it may well be that each epistemically possible scenario for the agent verifies $(A \wedge B)$ and $\neg A$. So we cannot use scenarios in the Single Disprovability Model to satisfy the rationality desideratum. To make progress on the rationality problem, scenarios need to obey further constraints than scenarios in the Single Disprovability Model.

Accordingly, we cannot use the Single Disprovability Model to satisfy the content and rationality desiderata. Since W_n, for any n, fails to contain only possible and non-trivially impossible scenarios, the Single Disprovability Model thus fails to successfully play the role of non-trivial epistemic space.

In response, one might firstly point out that scenarios in the Single Disprovability Model in a certain sense have the capacity to *represent* the falsity of jointly inconsistent sets of sentences.[8] For instance, though scenarios that contain $\{(A \wedge B), \neg A, \neg B\}$ are never eliminated from W_n, scenarios in W_n will still contain the sentence $C = \neg((A \wedge B) \wedge (\neg A \wedge \neg B))$, for sufficiently large n. Since C is provable in some number n of steps in $\mathcal{S}$, scenarios in W_n will at worst have as subsets sets of the form:

$$\Delta = \{\neg((A \wedge B) \wedge (\neg A \wedge \neg B)), (A \wedge B), \neg A, \neg B\}.$$

One can then attempt to let C represent in w the falsity of $\{(A \wedge B), \neg A, \neg B\}$. From this, one might try to extract the following two claims. First, since the epistemic n-intension of C is true at all $w \in W_n$, this represents the non-trivial inferential relations between $(A \wedge B)$ and A and B. Second, since C is true at all $w \in W_n$, this represents the idea that A and B should always be epistemically necessary for a moderately ideal agent when $(A \wedge B)$ is. Understood in this way, one might claim, there is a relevant sense in which the Single Disprovability Model can avoid the problems above.[9]

Whether or not we can make sense of this notion of representing the falsity of a jointly inconsistent set of sentences, we still cannot use the *construction*

[8] Thanks to Joe Salerno for discussion here.

[9] It is worth pointing out here that we should *not* amend the Single Disprovability Model with a clause that allows us to eliminate from W_n scenarios that contain sets of sentences whose conjunction is disprovable in n steps in $\mathcal{S}$. That would involve imposing a joint consistency constraint on scenarios, and the corresponding construction will face the problems of the Joint Disprovability Model that I will investigate shortly. For further details, see footnote 15, page 122.

of W_n to satisfy the content and rationality desiderata. We want to associate non-ideal epistemic intensions with *arbitrary* sentences, and not only with the class of n-provable sentences. Since scenarios in W_n can contain sets like Δ, it follows that the epistemic n-intensions are too ill-behaved to represent the contents of the epistemic states of moderately ideal agents. Similarly, we want to use the *construction* of W_n to infer facts about which sentences should remain epistemically possible for moderately ideal agents, given that they may accept thus and so. Since scenarios in W_n can contain sets like Δ, we cannot use the construction to infer that if $(A \wedge B)$ is true at all epistemically possible scenarios for a moderately ideal agent, then A is also true at all these scenarios. The fact that C is true at all epistemically possible scenarios for the agent does nothing to establish the required relations between $(A \wedge B)$ and A in these scenarios.

In response to the content and rationality problems, one might secondly point out that we can overcome these problems by "closing" scenarios under various inference rules. For instance, to overcome the content problem, we know that the n-intension of A must be true at a scenario $w \in W_n$, for some n, whenever the n-intension of $(A \wedge B)$ is true at w. In turn this means that A must be true at a $w \in W_n$ whenever $(A \wedge B)$ is true at w. To ensure that we capture such basic inferential relations among contingent sentences in the Single Disprovability Model, we might try to close scenarios in W_n under conjunction-elimination: For all sentences A and B and scenarios $w \in W_n$, if $(A \wedge B)$ is true at w, then A and B are also true at w. By closing scenarios under conjunction-elimination, it then follows immediately that the n-intension of A is true at a $w \in W_n$ whenever the n-intension of $(A \wedge B)$ is true at w. In a similar fashion, we can close scenarios in W_n under biconditional-elimination to capture the basic inferential relations among A, B, and $(A \leftrightarrow B)$. So given the way the content and rationality problems are stated above, we can overcome these problems by closing scenarios under conjunction-elimination and biconditional-elimination.

There are two problems with this strategy. First, it seems to be the wrong conceptual strategy for developing models of resource-bounded agents that can only engage in limited logical reasoning. By closing scenarios under conjunction-elimination, for instance, it follows that an agent that accepts a conjunction, however long and complicated, thereby also automatically accepts each of its conjuncts.[10] As a result, all logical reasoning with conjunction-elimination is cognitively cost-free, and this holds no matter how complex the relevant chain of logical reasoning may otherwise be. But logical reasoning is not cost-free for resource-bounded agents. So to model these agents, we should not close scenarios under inference rules such as conjunction-elimination—or at least we should only close scenarios in a restricted or limited way that parallels the limited computational capacities that moderately ideal agents have available for logical reasoning.

Second, it is obvious that we can restate the content and rationality problems in terms of the other standard inference rules. For instance, to satisfy the content desideratum, we need to ensure that the n-intension of B is true at a scenario $w \in W_n$ whenever A and $(A \rightarrow B)$ are true at w. Otherwise, we cannot represent the basic inferential relations among A, $(A \rightarrow B)$ and B in the corresponding n-intensions. For scenarios in W_n to do this job, B must be true at a $w \in W_n$ whenever A and $(A \rightarrow B)$ are true at w. But when A and B are contingent sentences, we cannot use scenarios in the Single Disprovability Model to establish such relations. An adherent of the model might reply to this problem by closing scenarios under modus ponens: For all sentences A and

[10] To see this, suppose an agent a accepts some complex conjunction $B = (A_1 \wedge A_2 \ldots \wedge A_n)$. Then B is epistemically necessary for a and hence true at all epistemically possible scenarios for a. If scenarios are closed under conjunction-elimination, then each conjunct A_i is also true at all epistemically possible scenarios for a. Then each A_i is epistemically necessary for a, and hence a accepts each A_i. So if a accepts B, a accepts each of its conjuncts A_i.

B and scenarios $w \in W_n$, if A and $(A \rightarrow B)$ are true at w, then B is also true at w. But now we can restate the content and rationality problems in terms of inference rules that involve $\vee$, and the adherent of the Single Disprovability Model can reply correspondingly by closing scenarios under the inference rules involving $\vee$. Clearly, this dialectics can keep going until all scenarios in W_n are closed under the standard inference rules.

But if all scenarios in W_n are closed under the standard inference rules, scenarios in W_n are closed under entailment simpliciter. But it then follows by familiar reasoning that all inconsistent, non-ideal scenarios in W_n are trivially impossible in the sense that they verify *all* sentences.[11] But to serve as a construction of a non-trivial epistemic space, only non-trivially impossible scenarios should survive in W_n. If so, scenarios that verify all sentences should be excluded from W_n because of their "explosive", "anything goes" nature. Yet, if we eliminate such "anything goes" scenarios from W_n, we thereby end up eliminating all inconsistent scenarios from W_n. As a result, only consistent, ideal scenarios survive in W_n, in which case W_n is only suitable for modeling ideal, logically omniscient agents. So to build models of non-trivial epistemic space that can overcome the content and rationality problems, we should not close scenarios under the standard inference rules.

In a certain sense, however, one might view most of the models of non-ideal epistemic space that I investigate in both this and the next chapter as attempting to close scenarios *restrictedly* or *limitedly* under the standard inference rules. The Joint Disprovability Model of the next section, for instance, requires that scenarios in non-trivial epistemic space cannot be jointly disproved by applying the standard inference rules a limited number n of times on sentences already verified by these scenarios. Roughly, the idea is that

[11] In fact, this problem arises as soon as scenarios contain all instantiations of the axiom schemas and are closed under modus ponens; see also footnote 15 on page 122.

whenever a scenario w verifies the premises of a given inference rule, then as long as w verifies all sentences that can be inferred from these premises by applying the rule n many times, w will survive in non-trivial epistemic space. In this sense, scenarios can be closed restrictedly or limitedly under the standard inference rules. So although we cannot solve the content and rationality problems by closing scenarios unrestrictedly under various inference rules, we might well attempt to do so by only closing them restrictedly or limitedly under such rules.

As Extreme Epistemic Space, we have seen that the Single Disprovability Model has its roles to play. It allows us to prove (C-$\Diamond_n$) and (C-$\Box_n$) and vindicate the (Epi-Pos) and (Epi-Nec) principles for the whole spectrum of level-n agents. But because arbitrarily jointly inconsistent scenarios need never be eliminated from W_n, the Single Disprovability Model fails to play the role of non-trivial epistemic space. In the next chapter, I will investigate versions of the Single Disprovability Model that aim to impose further constraints on scenarios in W_n. For now, however, I will investigate the most natural approach to tackle the limitations of the Single Disprovability Model.

4.2 The Joint Disprovability Model

The main limitations of the Single Disprovability Model arise because scenarios in any W_n can contain jointly inconsistent sets of sentences such as $\{(A \wedge B), \neg A, \neg B\}$. To overcome these limitations, the natural suggestion is to aim to eliminate scenarios that contain such sets from non-ideal epistemic space. To this end, I investigate the *Joint Disprovability Model* that only differs from the Single Disprovability Model in taking the basic notion to be that of *jointly disproving a set of sentences.*

I define the core notion of joint disprovability as follows:

> (**Joint Disprovability**) A set of sentences Γ is jointly disprovable within n steps in $\mathcal{S}$ iff a contradiction can be derived from Γ within n steps in $\mathcal{S}$.

Two comments on this definition. First, I allow for simplicity that $\mathcal{S}$ can make derivations from assumptions or premises, and as such, that we can eliminate sets of sentences by deriving a contradiction *from* such sets. In the Single Disprovability Model, by contrast, we can only eliminate those sets of sentences that contain a sentence the negation of which can be derived in $\mathcal{S}$ from no assumptions. Yet, we could say that a set of sentences Γ is jointly disprovable in n steps in $\mathcal{S}$ if and only if there is a conjunction of sentences in Γ the negation of which can be derived within n steps in $\mathcal{S}$ from no assumptions. For the general results to come, these differences are not important, but the results will be much easier to establish with (Joint Disprovability).[12] Second, for simplicity, I will think of contradictions as pairs $\{A, \neg A\}$. Yet we could easily introduce or define $\wedge$ and say that Γ is jointly disprovable within n steps in $\mathcal{S}$ whenever $(A \wedge \neg A)$ can be derived from Γ within n steps in $\mathcal{S}$. If we do this, we have to add one extra step to the results below.

We can now use (Joint Disprovability) to define a notion of joint consistency that applies to sets of sentences:

> (***n*-Consistency**) A set of sentences Γ is n-consistent with respect to $\mathcal{S}$ iff Γ cannot be jointly disproved within n steps in $\mathcal{S}$; otherwise Γ is n-inconsistent with respect to $\mathcal{S}$.

By taking scenarios to be minimally closed sets of sentences, we can then use (n-Consistency) to define the following spheres of scenarios, where V_J is the class of scenarios in the Joint Disprovability Model:

[12] For further details, see footnote 15, page 122.

The set V_0 of 0-consistent scenarios: the scenarios in V_J that cannot be jointly disproved in 0 steps in $\mathcal{S}$.

The set V_1 of 1-consistent scenarios: the scenarios in V_J that cannot be jointly disproved within 1 step in $\mathcal{S}$.

The set V_2 of 2-consistent scenarios: the scenarios in V_J that cannot be jointly disproved within 2 steps in $\mathcal{S}$.

$$\vdots$$

The set V_n of n-consistent scenarios: the scenarios in V_J that cannot be jointly disproved within n steps in $\mathcal{S}$.

$$\vdots$$

The set V_∞ of ∞-consistent scenarios: the scenarios in V_J that cannot be jointly disproved within any number of steps in $\mathcal{S}$.

I stipulate that no set of sentences can be jointly disproved in 0 steps in $\mathcal{S}$, in which case V_J is identical to the class of scenarios V_0. V_n corresponds to the class of scenarios in V_0 whose only constraint is that no contradiction can be derived from any scenario in the class within n steps in $\mathcal{S}$. When we increase the value for n, increasingly more non-ideal scenarios will be excluded from V_n. V_∞ corresponds to the class of scenarios from which no contradiction can be derived in any number of steps in $\mathcal{S}$.

As with scenarios in the Single Disprovability Model, scenarios in the Joint Disprovability Model are related by the subclass relation:

$$V_0 \supseteq V_1 \supseteq V_2 \ldots \supseteq V_n \ldots \supseteq V_\infty.$$

Intuitively, the idea is, as we move inwards through this epistemic space, it takes increasingly more computational effort to spot the contradictions that are verified by various n-inconsistent scenarios.

Though the Single Disprovability Model and the Joint Disprovability Model only differ in whether we take the core notion to be that of disproving a single sentence or jointly disproving a set of sentences, they give very different results and have very different problems. On the good side, the Joint Disprovability Model allows us to eliminate the trivially impossible scenarios that survive in the Single Disprovability Model. Take a scenario w such that $\{(A \wedge B), \neg A, \neg B\} \subseteq w$, where A and B are contingent sentences. No matter how large n goes, scenarios like w can survive in W_n in the Single Disprovability Model. But in the Joint Disprovability Model, w is quickly eliminated from V_n. Specifically, we can jointly disprove w in 2 steps in $\mathcal{S}$. First we apply conjunction-elimination on $(A \wedge B)$ to get A from w in 1 step. Second, we infer $\neg A$ from w in 1 step by use of the trivial inference rule that enables us to infer A in 1 step from any set Γ such that $A \in \Gamma$.[13] So A and $\neg A$ can be derived from w in 2 steps, and hence w is jointly disprovable in 2 steps. By (n-Consistency), w is then 2-inconsistent and fails to be in any V_n, for $n > 1$. So the Joint Disprovability Model constitutes an improvement in this respect. However, as we shall see now, this improvement is negligible compared to the problems that arise in the Joint Disprovability Model.

4.2.1 Problems in the Joint Disprovability Model

We want to use the sequence of epistemic spaces that we can isolate in the Joint Disprovability Model to establish the Carnap-style analyses (C-$\Diamond_n$) and (C-$\Box_n$). But we cannot use the sequence of epistemic spaces $\big(V_0, V_1, \dots\big)$ to establish these analyses. Let $\boxtimes_n$ be the kind of operator that receives the Carnap-style analysis in the Joint Disprovability Model:

(**C**-$\boxtimes_n$) $\boxtimes_n A$ iff for all scenarios $w \in V_n$, $A \in w$.

[13] For further clarification and motivation of why proof steps are counted in this way in systems that allow derivations from assumptions, see chapter 3, section 3.2, page 88.

Irrespective of what $\boxtimes_n$ means, it cannot mean 'provability in n steps in $\mathcal{S}$'. To see why, I will prove:

> (**Omni**) For all sentences B and $n > 1$, if B is provable in $\mathcal{S}$, then $\boxtimes_n B$.

Since it is not true that all logically true sentences can be proved in 2 steps in $\mathcal{S}$, (Omni) will entail that $\boxtimes_n$ cannot mean 'provability in n steps in $\mathcal{S}$'. And as a result of (Omni), the Joint Disprovability Model will fail as a model of non-trivial epistemic space.

To prove (Omni), I first prove (Disp2):

> (**Disp2**) For all sentences B, if B is provable in $\mathcal{S}$, then any maximal set of sentences Γ such that $\neg B \in \Gamma$ is jointly disprovable in 2 steps in $\mathcal{S}$.

For the purpose of establishing (Disp2), and hence (Omni), I restrict $\mathcal{S}$ to axiomatic systems in which every line of a proof is derived from premises, or in which every line is derived from the empty set if there are no premises. Although results similar to (Omni) can be established for other kinds of proof systems—as (G-Incon) in chapter 5 also shows—I can make my general point by focusing on axiomatic systems.

I prove (Disp2) by induction on the shortest number of steps required to disprove Γ:

> Base case: Assume B is provable in 1 step in $\mathcal{S}$ and that $\neg B \in \Gamma$—since no sentence is provable in 0 steps in $\mathcal{S}$, (Disp2) is vacuously true for $n = 0$, so the interesting base case is for $n = 1$. We want to show that Γ is jointly disprovable in 2 steps in $\mathcal{S}$. Since $\neg B \in \Gamma$, $\neg B$ can be derived from Γ in 1 step.[14] Since B is provable in 1 step,

[14] That is, $\neg B$ can be derived from Γ in 1 step in $\mathcal{S}$ by the trivial inference rule that enables us to infer A in 1 step from any set Γ such that $A \in \Gamma$.

B can be derived from any set in 1 step. So B can be derived from Γ in 1 step. So $\neg B$ and B can be derived from Γ in 2 steps, and hence Γ is jointly disprovable in 2 steps in $\mathcal{S}$. So (Disp2) holds for the base case.

Inductive step: Assume for the induction hypothesis that (Disp2) holds for all sentences B that are provable in n steps in $\mathcal{S}$. We want to show that (Disp2) holds when B is provable in $n+1$ steps in $\mathcal{S}$. If $n = 0$, we repeat the argument from the base case. For $n > 1$, suppose B is provable in $n+1$ steps in $\mathcal{S}$ and that $\neg B \in \Gamma$. Then B will have to be derived from certain assumptions $A_1, A_2, \ldots, A_k$ such that each A_i is provable in at most n steps and such that B can be derived from $A_1, A_2, \ldots, A_k$ in 1 step. To show that Γ is jointly disprovable in 2 steps in $\mathcal{S}$, there are two cases:

Case 1: For some A_i, $A_i \notin \Gamma$. Hence $\neg A_i \in \Gamma$ by (Maximality). By the induction hypothesis, for any A_i that is provable within in n steps in $\mathcal{S}$ and any set Γ, if $\neg A_i \in \Gamma$, then Γ is jointly disprovable in 2 steps in $\mathcal{S}$.

Case 2: For all A_i, $A_i \in \Gamma$. We now repeat the argument from the base case. Since B can be derived from $A_1, A_2, \ldots, A_k$ in 1 step, and since each $A_i \in \Gamma$, then B can be derived from Γ in 1 step. Since $\neg B \in \Gamma$, $\neg B$ can be derived from Γ in 1 step. So B and $\neg B$ can be derived from Γ in 2 steps, and hence Γ is jointly disprovable in 2 steps in $\mathcal{S}$.

So (Disp2) holds for the inductive step, and I conclude that (Disp2) holds in general. □

Because all scenarios in V_J are maximal, we then immediately get (Con2) by (Disp2):

(Con2) For all sentences B, if B is provable in $\mathcal{S}$, then B is true at each $w \in V_2$.

If B is provable in $\mathcal{S}$, we have by (Disp2) and (n-Consistency) that any scenario w such that $\neg B \in w$ is 2-inconsistent. So for all $w \in V_2$, $\neg B \notin w$. By (Maximality), then $B \in w$ for all $w \in V_2$. So (Con2) holds.

As a corollary to (Con2), we then get (Omni):

(Omni) For all sentences B and $n > 1$, if B is provable in $\mathcal{S}$, then $\boxtimes_n B$.

By (Con2), for any provable B in $\mathcal{S}$, $B \in w$ for all $w \in V_2$. By (C-$\boxtimes_n$), right to left, then $\boxtimes_2 B$. Accordingly, for all B and $n > 1$, if B is provable in $\mathcal{S}$, then $\boxtimes_n B$. So (Omni) holds.[15]

Given results like (Disp2) and (Omni), it is clear that the Joint Disprovability Model cannot play the role of non-trivial epistemic space. (Disp2) entails that all non-ideal scenarios are jointly disprovable in 2 steps in $\mathcal{S}$. Effectively, this means that (Disp2) collapses the intended stratified structure in the Joint Disprovability Model. Since all provable sentences are true at each scenario $w \in V_2$, V_2 ends up playing the role of V_∞ and all spaces in between are lost.

[15] If we decide on a broadly conjunctive definition of the notion of joint disprovability, it is worth pointing out an alternative proof of a result similar to (Omni). First, define (Joint Disprovability*) as: A set Γ is jointly disprovable within n steps in $\mathcal{S}$ iff there is a conjunction of sentences in Γ the negation of which can be derived within n steps in $\mathcal{S}$ from no assumptions. Second, assume all axiom schemas in $\mathcal{S}$ can be instantiated in 1 step in $\mathcal{S}$; the exact number matters little as long as it is sufficiently small, and in particular smaller than m below. Consider then $\neg(A \wedge ((A \rightarrow B) \wedge \neg B))$. For reasonable systems, $\neg(A \wedge ((A \rightarrow B) \wedge \neg B))$ is provable in some small m. By (Joint Disprovability*), then any w that contains $\{A, (A \rightarrow B), \neg B\}$ is jointly disprovable within m steps. So for all $w \in V_n$, for $n \geq m$, if $\{A, (A \rightarrow B)\} \subset w$, then $B \in w$. If all axiom schemas are instantiated in 1 step, and if modus ponens is a rule of inference in $\mathcal{S}$, then each $w \in V_m$ will contain all provable sentences. By (C-$\boxtimes_n$), we then have $\boxtimes_n B$ for all provable B, for $n \geq m$. Though m will not equal 2 in many systems, m will still be small. So we will have a result similar to (Omni): For all B and for $n \geq m$, if B is provable in $\mathcal{S}$, then $\boxtimes_n B$.

Apart from scenarios in the trivial epistemic spaces V_0 and V_1, where any provable sentence can be false, the stratified picture that motivates constructions of non-trivial epistemic space is lost in the Joint Disprovability Model.

(Omni) entails that any agent that can perform just 2 steps in $\mathcal{S}$ is logically omniscient. Given the role that we want operators like $\boxtimes_n$ to play in an analysis of deep$_n$ epistemic necessity, and given the definition of level-n agents, we can substitute $\Box_n$ with $\boxtimes_n$ in ($\Box_n$-Level-n) to get:

($\boxtimes_n$-Level-n) If $\boxtimes_n A$, then A is epistemically necessary for a level-n agent a_n.

By (Omni), then any logical truth is epistemically necessary for any level-2 agent. Plausibly, any moderately ideal agent can instantly perform at least 2 steps in $\mathcal{S}$. So (Omni) wrongly implies that all moderately ideal agents are characterized as logically omniscient. As a result, we cannot use the Joint Disprovability Model to capture facts about epistemic possibility for the class of moderately ideal agents that are logically competent, but not logically omniscient.

In the Joint Disprovability Model, we attempt to use the notion of joint disprovability to eliminate the trivially impossible scenarios that survive in the Single Disprovability Model. While the spaces V_0 and V_1 cannot help us achieve this goal, we have seen that the remaining epistemic space contains no impossible scenarios at all. So we cannot use the Joint Disprovability Model to make (Epi-Pos) and (Epi-Nec) plausible principles for the class of moderately ideal agents that have non-trivial but bounded cognitive capacities available for instant a priori reasoning. In contrast to the Single Disprovability Model, which pulls too much towards Extreme Epistemic Space, the Joint Disprovability Model pulls too much towards Ideal Epistemic Space.

4.3 Jago's Model of Epistemic Space

Models of non-ideal epistemic space that are similar in spirit to the Joint Disprovability Model are vulnerable to the same criticism. Recently, Jago has proposed a model of non-ideal epistemic space that bears much resemblance to the Joint Disprovability Model. According to Jago,

> [a] non-ideal epistemic space is one in which not all scenarios are maximally specific coherent ways the world might be. Some of these scenarios are not only metaphysically impossible but also impossible by the standards of classical logic: what is true according to such scenarios may be contradictory and need not be closed under classical consequence.[16]

Much like I do, Jago thinks of ideal epistemic space in broadly Chalmersian terms as the class of maximally specific a priori coherent ways the world might be. And much like I do, Jago thinks of extreme epistemic space in broadly Priestian terms as the class of open worlds.[17]

For much the same reasons that I gave in the criticism of Extreme Epistemic Space, Jago wants to avoid the trivial structure that emerges if non-ideal epistemic space can contain any old open world. He suggests that a non-trivial notion of epistemic possibility is tied to "what is expected of a sincere, rational (although not ideal) agent."[18] In particular, we expect such agents to reject *basic a priori impossibilities* or *basic a priori falsehoods*. Jago does not define the class of basic a priori impossibilities, but takes them "to be those that any competent language user would recognize as false, non-inferentially, on a priori grounds."[19] These basic a priori impossibilities include sentences such

[16] Jago (2009a): pp. 331-332.

[17] Open worlds, we remember, are akin to the kinds of "anything goes" scenarios that we find in Extreme Epistemic Space.

[18] Jago (2009a): p. 333.

[19] Jago (2009a): p. 334. Though Jago restricts his attention to agents that can *non-inferentially* reject such basic impossibilities, his model can easily be generalized to accommodate agents for whom this is not so.

as 'There are round squares', '0 = 1', 'It rains and it does not rain', and 'My bike is blue and red all over'.

Following a suggestion from Chalmers, Jago then takes a sentence A to be deeply epistemically possible when it is not obvious a priori that $\neg A$. To construct a non-ideal epistemic space that goes with this notion of deep epistemic possibility, Jago then attempts to

> [...] take all of Chalmers' a priori coherent worlds, plus some but not all open worlds, to underlie the epistemic scenarios. More precisely, I will take a centred open world to be an epistemic scenario iff no agent would fall below our epistemic expectations by believing anything that is true according to that world.[20]

An open world that fails to be an epistemic scenario corresponds roughly to what I call a trivially or blatantly inconsistent scenario. An open world that meets the standards for being an epistemic scenario corresponds roughly to what I call a non-trivially or subtly inconsistent scenario. For the fragment of a priori truths that are also truths of propositional logic, an a priori coherent world corresponds to what I call an ideal scenario.

Given this, Jago then aims to make the notion of epistemic expectations precise by placing a total order $\preceq$ on open and closed worlds: $w \preceq w'$ holds whenever our expectations are such that if we expect an agent to reject w' a priori, then we also expect the agent to reject w a priori.[21]

> The maximal elements with respect to $\preceq$ are the worlds which are coherent (in Chalmers' sense): for any such world w and all worlds w', $w' \preceq w$. The minimal elements with respect to $\preceq$ are those according to which some obvious a priori impossibility is true, where that impossibility is as basic as an a priori impossibility can be.[22]

[20] Jago (2009a): p. 333.

[21] To 'reject a world a priori' means 'to reject some truth according to the world a priori'; cf. Jago (2009a): p. 334.

[22] Jago (2009a): p. 334.

Intuitively, if we expect an agent to reject any coherent or closed world a priori, then we expect the agent to reject *all* worlds a priori, which is absurd for any world involving analysis of epistemic possibility. In contrast, let $\mathcal{X}$ be the set of basic a priori impossibilities, and let $|w|$ be the set of truths according to w. Then if $\mathcal{X} \cap |w| \neq \emptyset$, $w \preceq w'$, for any world w'. That is, the minimal elements with respect to $\preceq$ are all those worlds that verify a basic a priori impossibility and that we expect all minimally rational agents to reject a priori, if anything at all.

Now that we have the maximal and minimal elements with respect to $\preceq$, we need to decide when it holds in general. Jago's idea is as follows:

> In addition to expecting agents to treat each member of $\mathcal{X}$ as describing an epistemic impossibility, we also expect rational agents to perform basic inferences, in accordance with the meanings of 'and', 'or', 'for all' and so on. [...] Let $\mathcal{R}$ be a set of basic inference rules, such that we can expect any rational agent to have the ability to apply any of those rules (if its cognitive resources allow).[23]

Model theoretically, Jago interprets $\mathcal{R}$ as a binary relation $[[\mathcal{R}]]$ between worlds. In general, $(w, w') \in [[\mathcal{R}]]$ if and only if there is an instance of a rule schema $\{A_1, A_2, \ldots A_k\} \vdash B$ in $\mathcal{R}$ such that $\{A_1, A_2, \ldots A_k\} \subseteq |w|, B \notin |w|$, and $|w'| = |w| \cup \{B\}$. That is, if w verifies each premise A_i for a given rule in $\mathcal{R}$, then one application of that rule takes us to another w' that verifies everything w does plus the conclusion B, which can be obtained by applying the rule to the premises.

Let '$\Gamma \vdash_{\mathcal{R}} A$' abbreviate '$A$ is derivable from Γ using just the rules in $\mathcal{R}$'. Consider then two sets of sentences Γ and Δ, and suppose $\Gamma \vdash_{\mathcal{R}} A$ and $\Delta \vdash_{\mathcal{R}} B$, where $\{A, B\} \subseteq \mathcal{X}$. Furthermore, suppose that

> [...] the minimum number of inference steps required to obtain A from Γ is less than the minimum number of steps required to obtain B from Δ. Then

[23] Jago (2009a): pp. 334-335.

> there is an intuitive sense in which Γ is more obviously incoherent tha[n] Δ for, although both are incoherent, the incoherence of Δ is harder to spot (using $\mathcal{R}$) than the incoherence of Γ.[24]

Then, if we expect an agent to reject Δ as a description of an epistemic scenario, then we expect the agent to reject Γ too. We can then say that for two open worlds w and w' such that $|w| = \Gamma$ and $|w'| = \Delta$, $w \preceq w'$. It is in such cases that $w \preceq w'$ holds in general. Intuitively, if we expect an agent to reject w' a priori, then we also expect the agent to reject a priori any w that is more obviously incoherent than w'.

Jago's model $\mathcal{J}$ of epistemic space is then a tuple:

$$\left\langle W^C, W^O, \nu, \preceq \right\rangle$$

W^C and W^O are classes of closed and open worlds, ν is a propositional valuation function assigning a truth-value to each sentence at each world, and $\preceq$ is a total order on $W^C \cup W^O$.

To get the intended non-ideal epistemic space, Jago constrains $\preceq$ by $\mathcal{X}$ and $\mathcal{R}$. Let $[[\mathcal{X}]]$ be the class of open worlds such that $\mathcal{X} \cap |w| \neq \emptyset$. Jago then defines a function f from scenarios in W^O to $\mathbb{N}$ such that $f(w) = n$ if and only if:

(J1) there is a sequence $w_0 w_1, \ldots w_n$ of worlds in $W^C \cup W^O$ such that $(w_i, w_{i+1}) \in [[\mathcal{R}]]$ for each $i < n$, $w_0 = w, w_n \in [[\mathcal{X}]]$; and

(J2) there is no sequence $w_0 w_1, \ldots w_m$ with the properties from (J1) such that $m < n$.

(J1) gives the number n of inference steps required to derive a basic a priori impossibility from w using just the rules in $\mathcal{R}$. (J2) says that n is the smallest

[24] Jago (2009a): p. 335; I have changed 'ϕ' and 'ψ' to 'A' and 'B' throughout the quote.

number for which we can derive such a basic a priori impossibility from w. Intuitively, the cognitive resources required to reject w a priori is then a function of the number of inference steps that it takes to derive a basic impossibility from w.

At this point, we can easily use $\mathcal{J}$ to construct a stratified non-ideal epistemic space.[25] In particular, we can use the details of the model to define a notion of joint disprovability as follows:

> (**J-Joint Disprovability**) A set of sentences Γ is jointly disprovable within n steps iff $f(\Gamma) = n$.

That is, Γ is jointly disprovable within n steps just in case n is the smallest number of inference steps that are required to derive a basic a priori impossibility from Γ using the rules in $\mathcal{R}$. If we assume that no Γ is jointly disprovable in 0 steps, we can then use (J-Joint Disprovability) to construct a spherical epistemic space similar to the one in the Joint Disprovability Model.

4.3.1 Problems in Jago's Model of Epistemic Space

As seen, Jago's model is very similar to the Joint Disprovability Model. If so, it should inherit the problems of the Joint Disprovability Model. Since it is not entirely clear from Jago's presentation whether worlds in $\mathcal{J}$ are maximal or not, we can proceed in different ways.

We can specify a condition (C_1), according to which worlds in $\mathcal{J}$ are maximal in the sense that for any sentence A, either A is true at w or $\neg A$ is true at w. If $\mathcal{J}$ satisfies (C_1), then $\mathcal{J}$ is faulty for roughly the same reasons as the Joint Disprovability Model.

[25] Jago himself does not consider such a stratified construction. Instead, he investigates various borderline cases concerning which open worlds should not count as epistemic scenarios in addition to those in $[[\mathcal{X}]]$. By having a stratified construction of epistemic space, there is no principle reason for aiming to settle such questions.

We can specify a condition (C_2), according to which worlds in $\mathcal{J}$ can fail to be maximal but some non-maximal worlds can be ruled out using rules in $\mathcal{R}$. If $\mathcal{J}$ satisfies (C_2), then $\mathcal{J}$ is faulty for roughly the same reasons as the Joint Disprovability Model.

We can specify a condition (C_3), according to which worlds in $\mathcal{J}$ can fail to be maximal but no non-maximal world can be ruled out using rules in $\mathcal{R}$. If $\mathcal{J}$ satisfies (C_3), then though $\mathcal{J}$ is not faulty for the same reasons as the Joint Disprovability Model, Jago's general picture strongly indicates that $\mathcal{J}$ should not satisfy (C_3). I will initially motivate this latter claim.

First, if $\mathcal{J}$ satisfies (C_3), then it becomes very hard to see how the order $\preceq$ on worlds in $W^C \cup W^O$ is supposed to work. For instance, let $w^\star$ be a non-maximal world according to which only A and B are true, for compatible A and B, but according to which *all* other sentences are *indeterminate*. For any world w that we expect an ideal agent to reject a priori, we can ask whether $w^\star \preceq w$. If $w^\star \npreceq w$, this must mean that there is no way to reject $w^\star$ a priori using $\mathcal{R}$—otherwise, an ideal agent would have rejected $w^\star$ because it fails to verify arbitrarily many logical and mathematical truths. But then not all the maximal elements of $\preceq$ are Chalmers' ideal scenarios that verify every a priori true sentence. So we have good reason to rule out non-maximal worlds like $w^\star$ using $\mathcal{R}$. If so, there is good reason to hold that $\mathcal{J}$ should not satisfy (C_3).

Second, if $\mathcal{J}$ satisfies (C_3), then worlds like $w^\star$ intuitively fail to capture "genuine epistemic possibilities for rational, yet non-ideal, agents."[26] Plausibly, our epistemic expectations are such that we expect minimally rational agents not only to reject obvious a priori falsehoods but also to accept obvious a priori truths. To capture this central aspect of our epistemic expectations, we have reason to say that worlds like $w^\star$, at which obvious a priori truths such as $(A \rightarrow A)$ and $\neg(A \wedge \neg A)$ can be indeterminate should be rejected a priori

[26] Jago (2009a): p. 340.

using $\mathcal{R}$. If so, there is good reason to hold that $\mathcal{J}$ should not satisfy (C$_3$).

In fact, there is reason to hold that worlds in Jago's model are maximal and hence that $\mathcal{J}$ should neither satisfy (C$_3$) nor (C$_2$). First, Jago defines ν in $\mathcal{J}$ to be "a propositional valuation function, assigning a truth-value to each sentence at each world."[27] Since any normal propositional valuation function is two-valued, Jago presumably intends ν to be as well. If ν could have taken a third truth-value like *indeterminate*, one presumes that Jago would have specified the details. If so, there is good reason to hold that $\mathcal{J}$ should neither satisfy (C$_3$) nor (C$_2$).

Second, Jago says that we "expect rational agents to perform basic inferences, in accordance with the meanings of 'and', 'or', 'for all' and so on."[28] This strongly suggests that Jago conceives of rules in $\mathcal{R}$ as basic inference rules of classical logic, and hence that worlds can only be rejected a priori using classical rules. If so, there is good reason to hold that scenarios in $\mathcal{J}$ are maximal. For instance, if $|w^\star|$ is a world at which A and B are true, and if $w^\star$ is maximal such that '$(A \wedge B) \notin |w^\star|$' means that $\neg(A \wedge B)$ is true at $w^\star$, then $w^\star$ can be ruled out using just classical rules in $\mathcal{R}$. Since we can derive $(A \wedge B)$ from $|w^\star|$ by an application of conjunction-introduction, both $(A \wedge B)$ and $\neg(A \wedge B)$ turn out to be true at $w^\star$ when $w^\star$ is maximal. But if $w^\star$ is not maximal and if '$(A \wedge B) \notin |w^\star|$' does not mean that $\neg(A \wedge B)$ is true at $w^\star$, then classical rules will not allow us to rule out $w^\star$. Accordingly, since there is good reason to hold that $w^\star$ should be a priori rejected using $\mathcal{R}$ and that $\mathcal{R}$ consists of classical rules, there is hence good reason to hold that worlds in $\mathcal{J}$ are maximal. So there is good reason to hold that $\mathcal{J}$ should neither satisfy (C$_3$) nor (C$_2$).

I trust that the arguments above give us reason to say that $\mathcal{J}$ should not

[27] Jago (2009a): p. 335; see also Jago (2009a): footnote 13, p. 332.

[28] Jago (2009a): p. 335.

satisfy (C_3). If any old non-maximal world like $|w^\star| = \{A, B\}$ can count as an epistemic scenario, there are simply too many aspects of Jago's general picture that we cannot capture. And in any case, *if* $\mathcal{J}$ satisfies (C_3), we cannot use $\mathcal{J}$ to play the role of non-trivial epistemic space. For instance, if non-maximal worlds like $w^\star$ are never ruled out as epistemic scenarios, then we cannot use the resulting epistemic space to ensure that obvious logical truths always remain epistemically necessary for moderately ideal agents.[29] So I will set aside the possibility that $\mathcal{J}$ satisfies (C_3). Though there is also reason to hold that Jago's model should not satisfy (C_2) either, I will in due course assume that it does and show that $\mathcal{J}$ then inherits the problems of the Joint Disprovability Model. First, however, I will assume that $\mathcal{J}$ satisfies (C_1) and show that $\mathcal{J}$ then inherits the problems of the Joint Disprovability Model. In the end, I will then have shown that $\mathcal{J}$ cannot play the role of non-trivial epistemic space.

So I first assume that $\mathcal{J}$ satisfies (C_1) and hence that all worlds in $W^C \cup W^O$ are maximal. Then $\neg A$ is true at w whenever $A \notin |w|$. By construction, we know that a world w fails to be a priori coherent when a sentence A in $\mathcal{X}$ can be derived from $|w|$ using $\mathcal{R}$. For current purposes, I will focus on the class of logically incoherent worlds from which a contradiction $\{A, \neg A\} \in \mathcal{X}$ can be derived (simpliciter) using the rules in $\mathcal{R}$.[30]

Given this, I will now show that a contradiction can be derived from any logically incoherent world in just 1 step using the rules in $\mathcal{R}$. To this end, I prove (J-Disp), where '$|w| \vdash_{\mathcal{R}} B$' abbreviates '$B$ can be derived from the truths according to w using the rules in $\mathcal{R}$':

> (**J-Disp**) For all sentences B and all worlds w, if $B \notin |w|$ and $|w| \vdash_{\mathcal{R}} B$, then $f(w) = 1$.

[29] See also section 5.3, chapter 5.

[30] Strictly, this means that I allow pairs of sentences $\{A, \neg A\}$ to be in $\mathcal{X}$. If we are troubled with this, we can add an extra step for conjunction-introduction below.

The proof of (J-Disp) is analogous to the proof of (Disp2). First, I assume that nothing can be derived from $|w|$ in 0 steps. Second, given Jago's semantic encoding of $\mathcal{R}$ as a binary relation between worlds, I will not bother about the trivial inference rule that enables us to infer A in 1 step from any set Γ such that $A \in \Gamma$. If we want to represent this rule explicitly, (J-Disp) will hold for $f(w) = 2$ rather than $f(w) = 1$—as always, the exact value for n matters little for the general insight. Finally, let '$|w| \vdash^n_{\mathcal{R}} B$' abbreviate '$B$ can be derived from the truths according to w in n steps using the rules in $\mathcal{R}$', where one inference step corresponds to one application of a rule in $\mathcal{R}$.

I prove (J-Disp) by induction on the shortest number of steps required to derive a contradiction from $|w|$:

> Base case: Assume $B \notin |w|$ and $|w| \vdash^1_{\mathcal{R}} B$. Since $|w| \vdash^1_{\mathcal{R}} B$, there is then a w' such that $(w, w') \in [[\mathcal{R}]]$, where $|w'| = |w| \cup B$. So $B \in w'$. But also, since $B \notin |w|$, then $\neg B \in |w|$ and hence $\neg B \in |w'|$. So $w' \in [[\mathcal{X}]]$. So there is a sequence $w_0 w_1$ such that $(w_0, w_1) \in [[\mathcal{R}]]$, where $w_0 = w, w_1 = w'$ and $w_1 \in [[\mathcal{X}]]$. So $f(w) = 1$. So (J-Disp) holds for the base case.
>
> Inductive step: Assume for the induction hypothesis that (J-Disp) holds for all sentences B such that $|w| \vdash^n_{\mathcal{R}} B$. We want to show that (J-Disp) holds when $|w| \vdash^{n+1}_{\mathcal{R}} B$. If $n = 0$, we repeat the base case. So let $n > 1$, and assume $B \notin |w|$ and $|w| \vdash^{n+1}_{\mathcal{R}} B$. Then B has to be derived from certain assumptions $A_1, A_2, \ldots A_k$ such that each A_i can be derived from $|w|$ in at most n steps and such that B can be derived from $A_1, A_2, \ldots A_k$ in 1 step. To show that $f(w) = 1$, there are two cases:
>
> **Case 1**: For some A_i, $A_i \notin |w|$. By the induction hypothesis,
> if $A_i \notin |w|$ and $|w| \vdash^n_{\mathcal{R}} A_i$, then $f(w) = 1$.

Case 2: For all A_i, $A_i \in |w|$. We now repeat the argument from the base case. Since B can be derived from $A_1, A_2, \ldots, A_k$ in 1 step, and since each $A_i \in |w|$, then $|w| \vdash^1_{\mathcal{R}} B$. Since $|w| \vdash^1_{\mathcal{R}} B$, there is then a w' such that $(w, w') \in [[\mathcal{R}]]$, where $|w'| = |w| \cup B$. But also, since $B \notin |w|$, then $\neg B \in |w|$ and hence $\neg B \in |w'|$. So $w' \in [[\mathcal{X}]]$. So there is a sequence $w_0 w_1$ such that $(w_0, w_1) \in [[\mathcal{R}]]$, where $w_0 = w, w_1 = w'$ and $w_1 \in [[\mathcal{X}]]$. So $f(w) = 1$.

So (J-Disp) holds for the inductive step, and I conclude that (J-Disp) holds in general. □

(J-Disp) states that we can derive a contradiction from any logically incoherent world in just 1 step using the rules in $\mathcal{R}$. Following Jago, we assume that agents have a basic inferential ability to employ the rules in $\mathcal{R}$ in a priori reasoning. (J-Disp) then implies (Omni$_1$):

(**Omni**$_1$) Any agent that can perform just 1 step using the rules in $\mathcal{R}$ can a priori reject any logically incoherent world in $W^C \cup W^O$.

If an agent can reject a world w a priori, we can say that w is epistemically impossible for this agent. By (Omni$_1$), then all logically incoherent worlds are epistemically impossible for any agent that can perform just 1 inference step using the rules in $\mathcal{R}$. So any agent that can perform just 1 inference step is logically omniscient. Plausibly, any moderately ideal agent is able to perform at least 1 step using the rules in $\mathcal{R}$. But such agents should obviously not be characterized as logically omniscient. So if $\mathcal{J}$ satisfies (C$_1$), then (Omni$_1$) entails that $\mathcal{J}$ fails as a model for the class of moderately ideal agents that are logically competent, but not logically omniscient. As in the Joint Disprovability Model, agents that can engage in minimal, inferential a priori reasoning are wrongly characterized as logically omniscient. So if $\mathcal{J}$ satisfies (C_1), then $\mathcal{J}$ inherits the problems of the Joint Disprovability Model.

To avoid (J-Disp), worlds in $\mathcal{J}$ cannot be maximal. So we assume that $\mathcal{J}$ satisfies (C_2) and hence that worlds in $\mathcal{J}$ can fail to be maximal and that some non-maximal worlds can be ruled out using $\mathcal{R}$. First, when worlds can fail to be maximal, let us say that A is *true* at w just in case $A \in |w|$ and that A is *false* at w just in case $\neg A \in |w|$. Then we can say that A is *indeterminate* at w just in case $A \notin |w|$, where this means that neither $A \in |w|$ nor $\neg A \in |w|$. Let '$1(A)$', '$0(A)$', and '$\frac{1}{2}(A)$' respectively mean that A is true, A is false, and A is indeterminate, and let '$1(A)$ at w', '$0(A)$ at w', and '$\frac{1}{2}(A)$ at w' respectively mean that A is true, false and indeterminate at w.

Second, if worlds can fail to be maximal, we need to specify what it means to reject a non-maximal world a priori using $\mathcal{R}$. In some way or another, I take it, we have to utilize an idea along the following lines: if $\frac{1}{2}(A)$ at w, but either $1(A)$ or $0(A)$ can be derived from $|w|$ using $\mathcal{R}$, then w can be rejected a priori. When worlds are maximal, we can use $\mathcal{X}$ as above and say that w can be rejected a priori if a contradiction $\{1(A), 0(A)\}$ is derivable from $|w|$ using $\mathcal{R}$. But if worlds can be non-maximal, we need a broader specification. Given a rule-based model of epistemic space like Jago's, the most natural suggestion is to say that indeterminate and true or false sentences can stand in contradictory relationships; in fact, it is very hard to see which other feature of a rule-based model can be used to rule out non-maximal worlds. So we can stipulate that $\{1(A), \frac{1}{2}(A)\}$ and $\{0(A), \frac{1}{2}(A)\}$ in addition to $\{1(A), 0(A)\}$ count as contradictions. Call this enriched notion of a contradiction an *i-contradiction.*

We can then say that a world w in $W^C \cup W^O$ *fails to be logically coherent* just in case an i-contradiction can be derived (simpliciter) from $|w|$ using $\mathcal{R}$. If w is maximal, w fails to be logically coherent when w is logically incoherent in the sense that $\{1(A), 0(A)\}$ can be derived from $|w|$ using $\mathcal{R}$. If w is non-maximal, w fails to be logically coherent just in case either $\{1(A), \frac{1}{2}(A)\}$, $\{0(A), \frac{1}{2}(A)\}$ or $\{1(A), 0(A)\}$ can be derived from $|w|$ using the rules in $\mathcal{R}$. This requires that rules in $\mathcal{R}$ allow us to reason about indeterminacies. Except for the uninteresting constraint $(R^\star)$ below, I will leave the nature of such non-classical rules in $\mathcal{R}$ an open question and note that they are required if we want to rule out non-maximal worlds using *rules* in $\mathcal{R}$.

Along the lines above, we can then say that a world w can be rejected a priori when an i-contradiction can be derived from $|w|$ using $\mathcal{R}$. For the class of minimally rational agents, we can say that if an i-contradiction can be derived from $|w|$ in a few steps using the rules in $\mathcal{R}$, then w can be rejected a priori by a minimally rational agent. Intuitively, when a sentence A is indeterminate at w, but A or $\neg A$ can be easily or obviously derived from $|w|$ using $\mathcal{R}$, then w can be rejected a priori by a minimally rational agent. For instance, suppose $(A \rightarrow A) \notin |w|$ and $\neg(A \rightarrow A) \notin |w|$. Then w is non-maximal. We can plausibly expect that $(A \rightarrow A)$ can be derived from any set in few steps using the rules in $\mathcal{R}$. If so, then an i-contradiction $\{\frac{1}{2}(A \rightarrow A), 1(A \rightarrow A)\}$ can plausibly be derived from $|w|$ in few steps using $\mathcal{R}$. In such cases, non-maximal worlds can be rejected a priori by minimally rational agents.

The rough picture outlined above allows us to say that $\mathcal{J}$ satisfies (C_2). But it also immediately allows us to prove a result similar to (J-Disp). Assume it takes 1 step using the rules in $\mathcal{R}$ to infer $1(A)$ from any $|w|$ such that $A \in |w|$ and 1 step to infer $0(A)$ from any $|w|$ such that $\neg A \in |w|$. Let then m be the smallest number such that for any A:

$(R^\star)$ It takes m steps using a rule in $\mathcal{R}$ to derive $\frac{1}{2}(A)$ from any $|w|$ such that $A \notin |w|$ and $\neg A \notin |w|$.

Though matters are less clear when $\mathcal{R}$ can contain non-classical rules, we can stipulate the value for m in $(R^\star)$ to be small. In fact, insofar as many-valued reasoning is not fundamentally different from two-valued reasoning, we can also safely stipulate that $m = 1$ in $(R^\star)$. For now, however, I will leave the exact value for m imprecise but intuitively take m to be small. This will also serve to show the general reasoning behind results like (Disp2) and (J-Disp).

Given this, we can now show that an i-contradiction can be derived from any world that fails to be logically coherent in a few steps using the rules in $\mathcal{R}$. Omitting unnecessary formalism, I prove (J-Disp*) to this end:

> (**J-Disp***) For all sentences B and all worlds w, if $0(B)$ at w or $\frac{1}{2}(B)$ at w and $|w| \vdash_{\mathcal{R}} B$, then an i-contradiction can be derived from $|w|$ in $m + 2$ steps using $\mathcal{R}$.[31]

The proof strategy for (J-Disp*) is exactly the same as the one for (J-Disp). Since the reasoning behind (J-Disp) is familiar by now, I will simplify and only show a part of the inductive proof of (J-Disp*):

> Base case: Assume $\frac{1}{2}(A)$ at w and $|w| \vdash^1_{\mathcal{R}} B$. Then B can be derived from $|w|$ in 1 step, and hence $1(B)$ can be derived from $|w|$ in 2 steps. Since $\frac{1}{2}(B)$ at w, then neither $B \in |w|$ nor $\neg B \in |w|$. So we can derive $\frac{1}{2}(B)$ from $|w|$ in m steps. So we can derive an i-contradiction ($\{1(B), \frac{1}{2}(B)\}$) from $|w|$ in $m + 2$ steps using $\mathcal{R}$. Omitting the analogous reasoning for $0(A)$ at w, (J-Disp*) holds for the base case.[32]
>
> Inductive step: Assume for the induction hypothesis that (J-Disp*) holds for all sentences B such that $|w| \vdash^n_{\mathcal{R}} B$. We want to show

[31] As (J-Disp*) is stated, I omit reference to the part of Jago's formalism that concerns the function f. Also, I will omit reference to the part that concerns the binary relation $[[\mathcal{R}]]$ between worlds in the proof of (J-Disp*) below. But in both cases, nothing is lost with respect to the underlying meaning of the formalism.

[32] Given the stipulations above, when $0(A)$ at w and $|w| \vdash^1_{\mathcal{R}} B$, we can derive an i-contradiction ($\{1(B), 0(B)\}$) from $|w|$ in 3 steps using $\mathcal{R}$. As such, (J-Disp*) will hold for $m + 2 = 3$ for (at least) all *maximal* worlds.

that (J-Disp*) holds when $|w| \vdash_{\mathcal{R}}^{n+1} B$. If $n = 0$, we repeat the base case. So let $n > 1$, and assume $\frac{1}{2}(B)$ at w and $|w| \vdash_{\mathcal{R}}^{n+1} B$. Then B has to be derived from certain assumptions $A_1, A_2, \ldots A_k$ such that each A_i can be derived from $|w|$ in at most n steps and such that B can be derived from $A_1, A_2, \ldots A_k$ in 1 step. There are three cases to consider:

Case 1: For some A_i, $0(A_i)$ at w. By the induction hypothesis, if $0(A_i)$ at w and $|w| \vdash_{\mathcal{R}}^{n} A_i$, then an i-contradiction $(\{1(A_i), 0(A_i)\})$ can be derived from $|w|$ in $m+2$ steps using $\mathcal{R}$.

Case 2: For some A_i, $\frac{1}{2}(A_i)$ at w. By the induction hypothesis, if $\frac{1}{2}(A_i)$ at w and $|w| \vdash_{\mathcal{R}}^{n} A_i$, then an i-contradiction $(\{1(A_i), \frac{1}{2}(A_i)\})$ can be derived from $|w|$ in $m+2$ steps using $\mathcal{R}$.

Case 3: For all A_i, $1(A_i)$ at w. We now repeat the argument from the base case. Since B can be derived from $A_1, A_2, \ldots, A_k$ in 1 step using $\mathcal{R}$, and since each $A_i \in |w|$, then $|w| \vdash_{\mathcal{R}}^{1} B$. Then B can be derived from $|w|$ in 1 step, and hence $1(B)$ can be derived from $|w|$ in 2 steps. Since $\frac{1}{2}(B)$ at w, then neither $B \in |w|$ nor $\neg B \in |w|$. So we can derive $\frac{1}{2}(B)$ from $|w|$ in m steps. So we can derive an i-contradiction $(\{1(B), \frac{1}{2}(B)\})$ from $|w|$ in $m+2$ steps using $\mathcal{R}$.

Omitting the analogous reasoning for $0(A)$ at w, I conclude that (J-Disp*) holds for the inductive step and thus in general. □

(J-Disp*) states that we can derive an i-contradiction from any world that fails to be logically coherent in $m+2$ steps using $\mathcal{R}$. Since we can take m

in (J-Disp*) to be small, (J-Disp*) will inherit the force and consequences of (Disp2) and (J-Disp)—if $m = 1$ in ($R^\star$), (J-Disp*) will hold for all $n > 2$. (J-Disp*) implies (Omni$_2$):

> (**Omni**$_2$) Any agent that can perform $m + 2$ steps using the rules in $\mathcal{R}$ can a priori reject any world in $W^C \cup W^O$ that fails to be logically coherent.

As with (Omni) and (Omni$_1$), (Omni$_2$) entails that any agent that can perform $m + 2$ inference steps is logically omniscient. Since we can take m in ($R^\star$) to be small, we can hold that any moderately ideal agent is able to perform at least $m + 2$ steps using the rules in $\mathcal{R}$—if $m = 1$ in ($R^\star$), this is particularly plausible. So we end up with the wrong result that moderately ideal agents are characterized as logically omniscient. So if $\mathcal{J}$ satisfies (C$_2$), then (Omni$_2$) entails that $\mathcal{J}$ fails as a model for the class of moderately ideal agents that are logically competent, but not logically omniscient. So if $\mathcal{J}$ satisfies (C$_2$), then $\mathcal{J}$ inherits the problems of the Joint Disprovability Model.

More generally, the results above suggest that by merely including non-maximal scenarios in epistemic space, we do not avoid results like (Omni). Rather, one must also hold that non-maximal scenarios like $w^\star = \{A, B\}$ are *not* generally ruled out in a step-based fashion using inference rules in a system. In light of this, we are drawn towards accepting a model of epistemic space that satisfies condition (C$_3$), according to which no non-maximal scenario can be ruled out by step-based logical reasoning.[33] But as argued above, there are good reasons to say that models like Jago's should not satisfy (C$_3$). Very roughly, just as worlds that qualify as epistemic scenarios should not contain obvious a priori falsehoods, so worlds that qualify as epistemic scenarios should not fail to contain obvious a priori truths. So there is good reason to say that

[33] Or at least we must accept that *most* non-maximal scenarios are never eliminated from epistemic space; see section 5.3, chapter 5 for further discussion of these matters.

models like Jago's satisfy either (C_1) or (C_2), in which case we cannot use such models to capture facts about epistemic possibility for moderately ideal agents.

Jago attempts to develop a non-ideal epistemic space in which scenarios

> are subtly incoherent and so are not trivial or obvious impossibilities. This is why they correspond to genuine epistemic possibilities for rational, yet non-ideal, agents.[34]

But as we have seen, Jago's model threatens to imply that no logically incoherent scenarios remain epistemically possible for agents that can engage in minimal, rule-based a priori reasoning. As in the Joint Disprovability Model, logical omniscience sneaks in, and the resulting epistemic space pulls too much towards Ideal Epistemic Space.

4.4 Summary

Since we can establish (C-$\Diamond_n$) and (C-$\Box_n$) in the Single Disprovability Model, we can use the model to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for the whole spectrum of agents. Yet, in the absence of any joint consistency constraints on scenarios, the Single Disprovability Model yields too many trivially impossible scenarios, where almost any set of blatantly inconsistent sentences can be true. As a result, scenarios in the Single Disprovability Model cannot play the role of non-trivially impossible scenarios. In the Joint Disprovability Model and in Jago's model, we can eliminate the trivially impossible scenarios that survive in the Single Disprovability Model. Yet, in the presence of joint consistency constraints on scenarios, models like the Joint Disprovability Model yield too few impossible scenarios, where any logical falsehood can be true. As a result, we cannot use such models to en-

[34] Jago (2009a): p. 340.

sure that (Epi-Pos) and (Epi-Nec) are plausible principles for the broad class of moderately ideal agents.

I set out to investigate models of non-trivial epistemic space that fall between Extreme Epistemic Space and Ideal Epistemic Space. But generally, we have seen that models like the Single Disprovability Model pull too much towards Extreme Epistemic Space, whereas models like the Joint Disprovability Model pull too much towards Ideal Epistemic Space. A dilemma is emerging for constructions of non-trivial epistemic space, and I investigate the details in the next chapter.

Chapter 5

Dilemma for Models of Non-Trivial Epistemic Space

We have seen that attempts to construct non-trivial epistemic space threaten to yield either too many trivially impossible scenarios, where almost any set of blatantly inconsistent sentences can be true, or too few impossible scenarios, where any logical falsehood can be true. Given that scenarios are maximal, the following dilemma has been motivated, although not yet established:

> (**Dilemma**) Either scenarios contain joint blatant inconsistencies, in which case we have trivially impossible scenarios in epistemic space, or scenarios do not contain joint blatant inconsistencies, in which case we have logical omniscience.

In this chapter, I will first establish that the dilemma holds. To this end, I show that each maximal, logically inconsistent set of sentences contains a (joint) blatant inconsistency, and derivatively that all logically inconsistent scenarios are trivially impossible. When scenarios are maximal, this makes the dilemma vicious for the general idea behind constructions of non-trivial epistemic space. Scenarios either contain or do not contain a blatant inconsistency. If scenarios contain a blatant inconsistency, they are trivially impossible and the resulting space cannot play the role of non-trivial epistemic space. As we saw, this happens in models like the Single Disprovability Model. If scenarios do not contain a blatant inconsistency, then scenarios are logically consistent and the resulting space can only be used to model logically omniscient agents. As we saw, this happens in models like the Joint Disprovability Model. In either case, we are caught on the horns of a vicious dilemma.

For the purpose of this project, we must accept the first horn of the dilemma. If we want to do better than the Single Disprovabiliy Model, this leaves us with two options that I will discuss in the second part of this chapter. First, we allow that some, but not all trivially impossible scenarios may survive in non-ideal epistemic space. I will call the corresponding models of epistemic space *Intermediate Models*. Second, we allow that scenarios in non-ideal epistemic space may fail to be maximal. I will call the corresponding models of epistemic space *Partial Models*.

To stick as close as possible to the intuitive picture that I used in chapter 3 to motivate models of non-trivial epistemic space, I will take the Single Disprovability Model as the anchor of discussion. This has two benefits. First, the Single Disprovability Model can give us the basic Carnap-style analyses (C-$\Diamond_n$) and (C-$\Box_n$), which are required to ensure that (Epi-Pos) and (Epi-Nec) are plausible principles for the whole spectrum of agents. Second, to make progress on the content and rationality problems, we somehow need to do better than the Single Disprovability Model. To evaluate the prospects of Intermediate Models and Partial Models, we can then evaluate whether they manage to do that. I will argue that neither Intermediate Models nor Partial Models can successfully overcome the problems of the Single Disprovability Model, and hence that neither models can successfully play the role of non-trivial epistemic space.

By the end of this chapter, we will hopefully have good reason to say that successful constructions of non-ideal, yet non-trivial epistemic spaces are hard, if not impossible to find.

5.1 All Inconsistent Scenarios Are Trivially Impossible

To establish the dilemma, I will show that (Incon) holds when scenarios are maximal:[1]

> (**Incon**) All logically inconsistent scenarios are trivially impossible.

If we can show (Incon), we will have shown that trivially impossible scenarios must be admitted in epistemic space if we want to avoid logical omniscience.

To show (Incon), we need a few definitions. In particular, we need a definition of what it means for a scenario to be trivially impossible. Intuitively, scenarios are trivially impossible when they contain blatant inconsistencies, where blatant inconsistencies are the kinds of inconsistencies that moderately ideal agents can easily rule out a priori. To make this intuitive characterization precise, let us first define every instance of the following kinds of inconsistencies to be a (joint) *blatant inconsistency*:[2]

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

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

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

I trust that LNC-, MP-, and NC-inconsistencies plausibly deserve the label *blatant inconsistencies.* Regardless of what exactly we mean by 'blatant inconsistencies', instances of LNC-, MP-, and NC-inconsistencies are prime candidates.

[1] If a scenario w is maximal, we remember, then either A is true at w or $\neg A$ is true at w for all sentences A. By the account of truth and falsity, and the identification of scenarios with sets of sentences, this means in turn that either $A \in w$ or $\neg A \in w$.

[2] Note that this definition is not meant to be an exhaustive definition, but rather a minimal definition or characterization of what it means to be a joint blatant inconsistency.

First, on one very natural understanding of what it means to be a blatant inconsistency, it means to be an inconsistency that any minimally rational agent can easily rule out a priori. Irrespective of how precisely we represent the logical reasoning that minimally rational agents can engage in, it remains intuitively clear that they can perform the easy, obvious and feasible reasoning that suffices to rule out LNC-, MP-, and NC-inconsistencies. Maybe this is even more clearly seen if we rephrase the negated conditionals in NC-inconsistencies in terms of their truth-functionally equivalent conjunctions: $\{\neg A, (A \wedge \neg B), \neg B\}$, $\{\neg A, (A \wedge \neg B), B\}$, $\{A, (A \wedge \neg B), B\}$.[3] Formulated in this way, a simple application of conjunction-elimination will enable a given agent to infer a contradiction from these sets.

Second, every instance of a LNC-, MP-, and NC-inconsistency can be disproved in few steps in reasonable proof systems for propositional logic. In the case where $\mathcal{S}$ is a standard semantic tableaux system, any branch that contains an instance of a LNC-, MP-, or NC-inconsistency will close in 1 step. In the case where $\mathcal{S}$ is a natural deduction system, we can prove a contradiction from the conjunctive versions of LNC-, MP-, and NC-inconsistencies within 10 steps.[4] Though I will not go through the vast amount of proof systems and establish how many steps it exactly takes to disprove a LNC-, MP-, or NC-inconsistency in each such system, it is clear that this number will be small. Accordingly, since I encode a priori logical reasoning as steps in $\mathcal{S}$, we can then also ensure that any level-n agent, for small n, can easily come to reject each instance of a LNC-, MP-, and NC-inconsistency. So we can naturally line up

[3] Since the set $\{\neg, \wedge\}$ is adequate for expressing all truth-functions, we should note that nothing is lost with respect to the general picture if we define blatant inconsistencies in terms of these conjunctions.

[4] Of course, the exact number will depend on the exact inference rules of the particular natural deduction system. But in a slow Lemmon (1998) style natural deduction system, we can disprove each conjunctive version of a LNC-, MP-, and NC-inconsistency within 10 steps—and much quicker if we have rules that allow us to infer both conjuncts from a conjunction in one step.

the definition of blatant inconsistencies with the intuitive picture from chapter 3, in which I motivated the idea that blatant inconsistencies always remain epistemically impossible for moderately ideal agents.[5]

We can now define a scenario w to be *trivially impossible* if and only if w contains a blatant inconsistency. Since each instance of a LNC-, MP-, or NC-inconsistency can be easily ruled out a priori, the definition captures the intuitive idea that trivially impossible scenarios are the kinds of scenarios that any moderately ideal agent can easily rule out a priori.

Given these definitions and a definition of scenarios as maximal sets of sentences, we can then show (Incon) by showing (G-Incon):

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

As always, sets of sentences are formed over a language that has symbols $\neg$ and $\rightarrow$ that play the same inferential roles as classical negation and material implication. The other standard connectives are then treated as shorthand for their familiar definitions in terms of $\neg$ and $\rightarrow$—similar remarks apply to the proof of (Sat) below. So logically inconsistent sets such as $\{B, (A \wedge \neg B)\}$ do not make (G-Incon) false since they are treated as shorthand for $\{B, \neg(A \rightarrow B)\}$.

[5] To be sure, the sentences that may be involved in an instance of a LNC-, MP-, or NC-inconsistency can be of arbitrary finite length. In some sense, this might make us doubt that instances of LNC-, MP-, and NC-inconsistencies count as blatant across the scale. However, since it is not very plausible in general to distinguish between subtle and blatant inconsistencies by looking at the syntactical complexity of a sentence, I think this doubt is ultimately unfounded. First, measuring the length of a sentence does not provide for a plausible distinction between subtle and blatant inconsistencies: Compare '$\neg((\forall x(Fx \rightarrow Gx) \wedge \neg Gx) \rightarrow \neg Fx)$' with '$\exists x(\neg Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge (Fx \wedge Fx))))))))))$'. Second, counting the number of connectives does not provide for a plausible distinction: Compare '$((A \vee B) \wedge \neg A) \rightarrow \neg B$' with '$(\neg A \wedge (A \wedge (A \wedge (A \wedge (A \wedge (A \wedge (A \wedge A)))))))$'. Third, counting the number of different kinds of connectives does not provide for a plausible distinction: Compare '$\neg(\neg(A \rightarrow C) \rightarrow \neg((\neg A \rightarrow \neg B) \rightarrow \neg(\neg A \rightarrow \neg C)))$' with '$((A \vee B) \wedge \neg A) \rightarrow \neg B$'. Alternatively, in terms of length, connectives and quantifiers, the sentence 'Every even number ≥ 4 can be expressed as the sum of two primes' is simple. But I take it that neither it nor its negation is or should intuitively count as blatantly inconsistent. For at least one useful notion of blatant inconsistencies, it is the basic inferential relations among sentences and thoughts that matter, and this is the notion that I am trying to capture.

To prove (G-Incon), I will take a lead from one of the methods used to prove the completeness of propositional logic. First, let us say that a set of sentences Γ is *consistent* if and only if Γ is satisfiable, where Γ is *satisfiable* if and only if there is an evaluation that makes each sentence in Γ true. Second, let $\mathcal{I}$ be an interpretation function that assigns either true or false, but not both, to each atomic sentence in our languages. Then ν is the following (classical) evaluation function:

($\nu\mathcal{I}$) If A is an atomic sentence, 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 $\nu(B) = T$.

This notion of consistency captures the idea that possible scenarios correspond to maximal sets of sentences that have a model or an interpretation in propositional logic.[6] The semantic clauses for $\neg$ and $\rightarrow$ capture the idea that the languages under consideration have symbols that play the same inferential roles as classical negation and material implication.

Given this, we can now prove (Sat) and derivatively use (Sat) to prove (G-Incon):

(**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 $B \in \Gamma$.

The proof of (Sat) goes as follows:

Let Γ be any set of sentences that satisfies (i) and (ii). We want to show that there is an evaluation function ν that makes each $A \in \Gamma$

[6] For broadly linguistic constructions of logically possible worlds, this minimal model-theoretical notion of consistency underlies standard definitions of worlds as maximal sets of sentences that "can all be true together". Of course, this is only a minimal specification, since most ersatz constructions of possible worlds need more than propositional logic and more than mere logical consistency.

true. To this end, we stipulate an interpretation $\mathcal{I}$ such that for all atomic A:

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

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

This is a possible stipulation, since $\mathcal{I}$ 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 of symbols it contains:[7]

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 $\nu(B) = T$.

[7] For instance, $\neg A$ is longer than A, and $(A \rightarrow B)$ is longer than A and B. Also, brackets do not count.

> 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 the proof of (Sat).[8] □

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

We can now use (Sat) to prove (G-Incon):

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

The proof of (G-Incon) goes as follows:

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 triples or pairs 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 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 $\{\neg A, \neg(A \rightarrow$

[8] Similar results hold for the other basic connectives of propositional logic. Assume, for instance, that our languages have a symbol $\wedge$ that plays the role of classical conjunction. Assume $(A \wedge B) \in \Gamma$, for some maximal set of sentences Γ that satisfies (i), (ii) and (iii): $(A \wedge B) \in \Gamma$ iff $A \in \Gamma$ and $B \in \Gamma$. If $(A \wedge B) \in \Gamma$, we want to show that $\nu(A \wedge B) = T$, where $\nu(A \wedge B) = T$ iff $\nu(A) = T$ and $\nu(B) = T$. By (iii), if $(A \wedge B) \in \Gamma$, then $A \in \Gamma$ and $B \in \Gamma$. By induction hypothesis, since A and B are shorter than $(A \wedge B)$, $\nu(A) = T$ and $\nu(B) = T$. Hence $\nu(A \wedge B) = T$.

$B), \neg B\}$, $\{\neg A, \neg(A \to B), B\}$, or $\{A, \neg(A \to B), B\}$. That is, Δ contains either a MP-inconsistency or a NC-inconsistency.

Derivatively, all maximal, logically inconsistent set of sentences contain a LNC-, MP-, or NC-inconsistency. So (G-Incon) holds. □

Given (G-Incon), (Incon) now follows by the definition of a trivially impossible scenario:

(**Incon**) All logically inconsistent scenarios are trivially impossible.

So when scenarios are maximal, we have proved (Incon).

When scenarios are maximal, (Incon) thus establishes (Dilemma) to the effect that if logical omniscience fails, then epistemic space must contain trivially impossible scenarios. So when scenarios are maximal, we can conclude that the general idea behind models of non-trivial epistemic space cannot be made precise: There is no non-ideal epistemic space that contains only possible and non-trivially impossible scenarios. More generally, when scenarios are maximal, we must accept the Scylla of trivially impossible "anything goes" scenarios to avoid the Charybdis of logical omniscience.

For projects similar to mine, we must accept the first horn of the dilemma. This leaves us with two options for doing better than the Single Disprovability Model: *Intermediate Models* and *Partial Models*. First, I investigate Intermediate Models that eliminate some, but not all trivially impossible scenarios from non-ideal epistemic space. Second, I investigate Partial Models that allow that scenarios in non-ideal epistemic space may fail to be maximal. I argue that neither Intermediate Models nor Partial Models can successfully overcome the problems of the simple Single Disprovability Model, and hence that neither models can play the role of non-trivial epistemic space.

5.2 Intermediate Models

To investigate Intermediate Models, I will use the core features of the Single Disprovability Model. First, a scenario w is identified with an arbitrary set of sentences in $\mathcal{L}$—that is, with an arbitrary set of possible sentence types in English*. Second, a sentence A is true at scenario w if and only if $A \in w$, and A is false at w if and only if $A \notin w$. Third, scenarios are minimally closed in the sense that for all A, $A \in w$ if and only if $\neg A \notin w$. So scenarios are maximal in the sense that for all A and w, either A is true at w or $\neg A$ is true at w. Fourth, W_n is defined as the class of scenarios that contain all sentences that are provable within n steps in $\mathcal{S}$. Formally, $W_n = \{w \in W_0 | PR_n \subset w\}$, where PR_n is the set of n-provable sentences in $\mathcal{S}$. The resulting epistemic space then corresponds to the sequence of epistemic spaces $\big(W_0, W_1, \dots\big)$.

Intermediate Models aim to improve on the Single Disprovability Model by invoking an additional constraint that scenarios need to satisfy to be in W_n. The primary job of this constraint is to eliminate some, but not all trivially impossible scenarios from W_n. There are two questions to ask. First, how can we motivate such a constraint? Second, how do we define it?

Though the motivational part is tricky, I take it that there is an intuition according to which Δ is more ill-behaved than Γ:

$$\Delta = PR_n \cup \{A, (A \rightarrow B), \neg B, C, (C \rightarrow D), \neg D\}.$$

$$\Gamma = PR_n \cup \{A, (A \rightarrow B), B, C, (C \rightarrow D), \neg D\}.$$

Whereas Δ contains two blatant inconsistencies, Γ contains only one. As a rough gloss, if I presented you with Δ as a hypothesis about a specific way things might be, it would offer you two ways of reasoning to an absurdity, whereas Γ only offers one. The inconsistency of Δ, so to speak, manifests itself more explicitly than the blatant inconsistency of Γ. The reason, one might intuit, why some but not all trivially impossible scenarios survive in epistemic space is simply that some of them contain *too many* blatant inconsistencies to be live options for the broad class of agents that are not extremely non-ideal.

Admittedly, this is weak motivation. But if Intermediate Models improve significantly on the Single Disprovability Model, we can always look for better motivation. So as a first pass, I will say that Intermediate Models require not only that scenarios in W_n contain all n-provable sentences, but also that they cannot contain *too many* blatant inconsistencies. To improve minimally on the Single Disprovability Model, Intermediate Models then have to ensure that scenarios in W_n cannot contain arbitrarily many blatant inconsistencies.

We can now invoke a notion of *independent inconsistencies* and say that two inconsistencies are *independent* just in case they do not have an atomic sentence in common. For instance, the inconsistency $\{A, (A \rightarrow B), \neg B\}$ is independent of the inconsistency $\{C, (C \rightarrow D), \neg D\}$, where A through D are distinct atomic sentences. Since we cannot (interestingly) distinguish between inconsistent scenarios by counting steps in a proof system, the idea is to distinguish between inconsistent scenarios by counting the number of independent inconsistencies that they contain. On this way of counting, an inconsistent scenario, all of whose inconsistencies involve atomic A contains only 1 independent inconsistency. If we were to remove all sentences that contain A from the scenario, the corresponding set of sentences would be satisfiable and hence consistent. An inconsistent scenario, some of whose inconsistencies involve atomic A but not B, and some of whose inconsistencies involve atomic B but not A, will contain 2 independent inconsistencies. If we were to remove all sentences that contain A and B from the scenario, the scenario would be consistent—and so on and so forth.[9]

[9] Another counting strategy begins with the class of maximally consistent sets of sentences. By adding and removing single sentences from a maximally consistent set Γ, we can intuitively increase and decrease the degree of inconsistency of Γ. For instance, assume

Given this counting method, we can now investigate the details of a constraint that secures that scenarios in W_n do not contain too many independent inconsistencies. I will investigate two ways of defining this constraint, and correspondingly distinguish between two versions of Intermediate Models. First, we have *Intermediate Models* $\mathcal{A}$ that aim to ensure that each *individual* scenario in W_n contains at most finitely many independent inconsistencies. We can see Intermediate Models $\mathcal{A}$ as aiming to exclude from non-ideal epistemic space each trivially impossible scenario that contains more than a certain finite number of blatant inconsistencies. Second, we have *Intermediate Models* $\mathcal{B}$ that aim to ensure that the *class* of scenarios W_n at most contains finitely many independent inconsistencies. We can see Intermediate Models $\mathcal{B}$ as aiming to exclude from non-ideal epistemic space any trivially impossible scenario that contains certain kinds of blatant inconsistencies. I investigate each version of an Intermediate Model in turn.

5.2.1 Intermediate Models $\mathcal{A}$

To develop an Intermediate Model $\mathcal{A}$, we aim to ensure that each individual scenario in W_n at most contains finitely many independent inconsistencies. We can define the following constraint on scenarios in W_n:

> (**Int-**$\mathcal{A}$) For any finite n and m greater than 0, if $w \in W_n$, then w contains at most m independent inconsistencies.

As in the unamended Single Disprovability Model, no scenario is excluded from W_0. When n is arbitrarily large, we can stipulate that each scenario $w \in W_\infty$ contains 0 independent inconsistencies. Hence scenarios in W_∞ are logically

$A \in \Gamma$. Assume we add $\neg A$ to Γ to make $\Gamma \cup \neg A$ inconsistent. Since Γ is maximally consistent, $\{A, \neg\neg A, \neg\neg\neg\neg A, \ldots\} \subset \Gamma$. By adding $\neg A$ to Γ, we then create infinitely many new inconsistencies, some of which are the pairs $\{A, \neg A\}, \{\neg\neg A, \neg A\}, \{\neg\neg\neg\neg A, \neg A\}, \ldots$. By removing, say, $\neg\neg A$ from $\Gamma \cup \neg A$, we do not restore the consistency of Γ. But if we remove $\neg A$, we do. We could then start counting these inconsistency-generating problem sentences. What I say below will apply to this counting strategy as well.

consistent. For intermediate W_n, each scenario in W_n can at most contain m independent inconsistencies.

Though it will not matter for current purposes, it is worth noticing the following. Suppose we stipulate that the *total* number l of independent inconsistencies is finite.[10] We could then define a limited stratified structure by saying that for any finite n and l such that $n < l$, if $w \in W_n$, then w contains at most $(l - n)$ independent inconsistencies. For example, if the total number of independent inconsistencies is 40, then any $w \in W_{10}$ contains at most 30 independent inconsistencies, any $w \in W_{11}$ contains at most 29 independent inconsistencies, and so on. On this version, scenarios in W_0 can contain any finite number of independent inconsistencies, and we can again stipulate that scenarios in W_∞ contain 0 independent inconsistencies. We can then intuitively say that scenarios in W_n contain more and more n-provable sentences and less and less independent inconsistencies as n increases (to a certain limit). To ensure the (C-$\Diamond_n$) and (C-$\Box_n$) analyses, it is of course important that n is smaller than l. That is, since l is finite, we have to stop eliminating scenarios from $W_{l-(l-1)}$ by decreasing the number of independent inconsistencies that scenarios in this space may contain. If we did not, we would eliminate all independent inconsistencies from each scenario in W_l, and the resulting class of scenarios would be logically consistent. Derivatively, for any A that is provable in more than l steps, there would not be enough scenarios in W_l to ensure that $\Diamond_l \neg A$ and $\neg \Box_l A$. This would violate the (C-$\Diamond_n$) and (C-$\Box_n$) analyses.

In contrast to the unamended Single Disprovability Model, we now have a constraint that can prevent inconsistent scenarios in W_n from being arbitrarily jointly inconsistent. Given the construction, each $w \in W_n$ can of course still be arbitrarily jointly inconsistent with respect to some finite number of

[10] If we stipulate that there are only finitely many independent inconsistencies, then on the current definition we stipulate that there are only finitely many distinct atomic sentences.

independent inconsistencies. But at least we can eliminate some of the trivially impossible scenarios that survive in the Single Disprovability Model. The question is now whether this is enough to make progress on the content and rationality problems. I think not. To make the discussion maximally vivid, I will assume that each scenario $w \in W_n$, for $n > 0$, at most contains 1 independent inconsistency.

For the content desideratum, we want to use non-ideal epistemic intensions defined over scenarios in W_n to represent the contents of the epistemic states of moderately ideal agents. Let the *epistemic n-intension* of a sentence be a function from scenarios in W_n to a truth-value. When $n > 0$, W_n is the class of scenarios each of which contains all n-provable sentences and at most 1 independent inconsistency. As in the Single Disprovability Model, we are immediately guaranteed that the epistemic n-intension of any n-provable sentence is necessary. But when we look beyond the class of n-provable sentences, Intermediate Models $\mathcal{A}$ inherit the limitations of the Single Disprovability Model.

To see this, let C and D be any two contingent sentences such that C contains the atomic sentence A and such that D contains the atomic sentence B. Suppose $(C \leftrightarrow D)$ is provable within n steps in $\mathcal{S}$. To capture the basic inferential relations between $(C \leftrightarrow D)$ and C and D, we want the n-intensions of C and D to coincide in truth-value when the n-intension of $(C \leftrightarrow D)$ is necessary. But we cannot use Intermediate Models $\mathcal{A}$ to ensure this: Though the n-intension of $(C \leftrightarrow D)$ is true at all scenarios in W_n, the values of the n-intensions of C and D can come arbitrarily apart at many scenarios in W_n. Since all independent inconsistencies remain present throughout the space W_n, there are plenty of scenarios in W_n all of whose inconsistencies involve A, and there are plenty of scenarios in W_n all of whose inconsistencies involve B. As a result, we cannot ensure that the values of the n-intensions of C and D coincide whenever the n-intension of $(C \leftrightarrow D)$ is necessary. Hence Intermediate Models

$\mathcal{A}$ do not allow us to capture basic inferential relations among thoughts and sentences, and consequently we cannot use these models to make progress on the content problem. So we cannot use Intermediate Models $\mathcal{A}$ to satisfy the content desideratum.

For the rationality desideratum, we want to use scenarios in W_n to give a world involving analysis of a non-trivial notion of epistemic possibility that captures which sentences should remain epistemically possible for minimally rational agents. We cannot use Intermediate Models $\mathcal{A}$ to satisfy the rationality desideratum. To see why, consider any two contingent atomic sentences A and B.[11] Assume that a minimally rational agent a accepts $(A \wedge B)$. Since a can easily come to infer A from $(A \wedge B)$, she rationally should accept A when she accepts $(A \wedge B)$. We capture this normative component in the (EP) analysis of epistemic possibility by saying that if $(A \wedge B)$ is epistemically necessary for a, then so is A. To analyze this non-trivial notion of epistemic possibility at the level of scenarios, we then need to ensure that A is true at all epistemically possible scenarios for a when $(A \wedge B)$ is. Intermediate Models $\mathcal{A}$ cannot do this job: For any W_n, though $(A \wedge B)$ is true at each $w \in W_n$ that remains epistemically possible for a, there is no guarantee that A is also true at each such w. Since all independent inconsistencies are present throughout W_n, the construction allows that scenarios w such that $\{(A \wedge B), \neg A\} \subset w$ may remain epistemically possible for a. Thus we cannot use scenarios in W_n to infer basic claims about which sentences should remain epistemically possible for minimally rational agents. So we cannot use Intermediate Models $\mathcal{A}$ to satisfy the rationality desideratum.

So with respect to the content and rationality desiderata, Intermediate Models $\mathcal{A}$ fare no better than the Single Disprovability Model. In fact, there

[11] For the general points, it does not matter that A and B are atomic sentences, but it facilitates the discussion.

might be reason to say that Intermediate Models $\mathcal{A}$ fare worse than the Single Disprovability Model. To motivate this, assume again that each inconsistent scenario in W_n, for $n > 0$, can contain at most 1 independent inconsistency. Consider then any level-n agent a_n, for $n > 0$, that accepts A, $(A \rightarrow B)$ and $\neg B$, where B is atomic. Then $\{A, (A \rightarrow B), \neg B\} \subset w$ for each $w \in W_n$, for $n > 0$, that remains epistemically possible for a_n. Given this, it now follows that a_n is strikingly close to being logically omniscient.

To see why, remember first that (G-Incon) says that all maximal, logically inconsistent sets of sentences contain a LNC-, MP-, or NC-inconsistency. We can then consider any MP- or NC-inconsistency that does not involve B and ask whether there is any epistemically possible scenario for a_n that contains such an inconsistency.[12] Without loss of generality, suppose there is an epistemically possible scenario for a_n that verifies the MP-inconsistency $\{C, (C \rightarrow D), \neg D\}$, where C and D do not involve B. There is then a scenario $w \in W_n$ such that w is epistemically possible for a_n and such that $\{A, (A \rightarrow B), \neg B, C, (C \rightarrow D), \neg D\} \subset w$. But since w contains two independent inconsistencies, w cannot be in W_n. So the MP-inconsistency $\{C, (C \rightarrow D), \neg D\}$ cannot be contained in any w that is epistemically possible for a_n. Thus, more generally, every epistemically possible scenario for a_n cannot contain a MP- or NC-inconsistency that does not already involve B. Since scenarios are maximal, it then follows by (G-Incon) that any scenario w that remains epistemically possible for a_n is consistent with respect to any sentence C and any set of sentences Γ that do not involve B.

Assume next that all axiom schemas are instantiated in 1 step in $\mathcal{S}$. By construction of W_n, each $w \in W_n$, for $n > 0$, then contains all instances of the axiom schemas in $\mathcal{S}$. If we assume further that $\mathcal{S}$ has modus ponens as its sole

[12] Since scenarios in Intermediate Models are minimally closed, they are also minimally consistent. As such, they never verify any instance of a LNC-inconsistency. So for the current argument, we do not have to consider LNC-inconsistencies.

rule of inference, then any provable C follows from the axioms by repeated applications of modus ponens.[13] So, since there are no epistemically possible scenarios for a_n that contain a MP- or NC-inconsistency that does not involve B, we get that all $w \in W_n$ that remain epistemically possible for a_n contain every logical truth C that does not involve B and whose proof does not depend on any sentence that involves B—to save breath, any logical truth C that does not *depend* on B. Derivatively, we get that any level-n agent that accepts each member of a blatantly inconsistent set of sentences such as $\{A, (A \rightarrow B), \neg B\}$ will be logically omniscient with respect to any logical truth C that does not depend on B.

But structurally and intuitively, this should not follow. An agent that accepts a blatantly inconsistent set of sentences such as $\{A, (A \rightarrow B), \neg B\}$ should not thereby be described as logically omniscient with respect to any logical truth C that does not depend on the atomic B. Of course, if we allow each scenario in W_n to contain more than 1 independent inconsistency, we can avoid this particular problem. But as long as scenarios in W_n can contain at most m independent inconsistencies, for some finite m, the structurally awkward feature of Intermediate Models $\mathcal{A}$ remains: Any agent that accepts m-many different kinds of blatant inconsistencies will be logically omniscient with respect to any logical truth C that does not depend on any of the atomic sentences in the m-many independent inconsistencies.[14] Intuitively, agents are not epistemically punished but rather rewarded for accepting blatant inconsistencies.

So it remains clear that we cannot use Intermediate Models $\mathcal{A}$ to satisfy the content and rationality desiderata, and hence that we cannot use Intermediate

[13] As always, we can easily generalize the reasoning to systems that have more inference rules. And as always, it does not matter for the general insight that all axiom schemas are instantiated in exactly 1 step, but it facilitates the discussion.

[14] Two blatant inconsistencies are of a 'different kind' when they do not have an atomic sentence in common.

Models $\mathcal{A}$ to play the role of non-trivial epistemic space. The question is then whether Intermediate Models $\mathcal{B}$ can do better.

5.2.2 Intermediate Models $\mathcal{B}$

To develop an Intermediate Model $\mathcal{B}$, we aim to associate a maximal number of independent inconsistencies with *spheres* of scenarios. More precisely, we can define the following:

> (**Level-m Inconsistency**) For any finite n and m greater than 0, W_n has level$-m$ iff the maximal number of independent inconsistencies contained by the class of scenarios in W_n is m.

For instance, if W_n has level-15, then any scenario that contains more than 15 independent inconsistencies is excluded from W_n. And if there are 16 scenarios, each of which contains a set of inconsistencies that involve a different atomic sentence, then one of them will not survive in W_n. For Intermediate Models $\mathcal{B}$, it does not matter whether these m independent inconsistencies are contained by a particular scenario in W_n, by all inconsistent scenarios in W_n, or arbitrarily distributed across inconsistent scenarios in W_n. Rather, it matters that only independent inconsistencies of particular kinds can survive in the class W_n—for instance all those inconsistencies that involve either one of the following 15 atomic sentences $A_1, A_2, \ldots, A_{15}$.

We can define the following constraint on W_n:

> (**Int-$\mathcal{B}$**) For some finite n and m greater than 0, W_n has level-m.

As above, no scenario is excluded from W_0, and we can stipulate that W_∞ has level-0, in which case all scenarios in W_∞ are logically consistent. As above, we can easily change the details of (Int-$\mathcal{B}$) to gain more flexibility. For instance, if we stipulate that there are only finitely many independent inconsistencies, we can develop a limited stratified space by saying that for any finite n and l such that $n < l$, W_n has level-$(l - n)$.

But Intermediate Models $\mathcal{B}$ immediately run into problems. In particular, the kinds of omniscience worries that could arise in Intermediate Models $\mathcal{A}$ are bound to arise in Intermediate Models $\mathcal{B}$. This is easily seen by the line of reasoning from above:

Suppose W_n has level-m, for some finite n and m greater than 0. Let the relevant set of inconsistencies all involve one of the atomic sentences $A_1, A_2, \ldots, A_m$. Consider then any MP- or NC-inconsistency that does not involve any of the atomics A_i. Without loss of generality, suppose $\{B, (B \rightarrow C), \neg C\} \subset w$, for some $w \in W_n$, where B and C do not involve any A_i. But when W_n has level-m, and when B and C do not already involve an A_i, this is not possible. So there is no scenario $w \in W_n$ that contains a MP- or NC-inconsistency that does not involve an A_i. By (G-Incon), any $w \in W_n$ is thus consistent with respect to any sentence B and any set of sentences Γ that do not involve any A_i. Assume then that all axiom schemas are instantiated in 1 step in $\mathcal{S}$, and that $\mathcal{S}$ has modus ponens as its sole rule of inference. By construction of W_n, each $w \in W_n$, for $n > 0$, then contains all instances of the axiom schemas in $\mathcal{S}$. Since any provable B follows from the axioms by repeated applications of modus ponens, and since no $w \in W_n$ contains a MP- or NC-inconsistency that does not involve an A_i, we get that all $w \in W_n$ contain every logical truth B that does not depend on any A_i. By the Carnap-style analysis (C-$\Box_n$), we have $\Box_n B$ if and only if $B \in w$ for all $w \in W_n$. So we get (Omni$_3$) immediately:

(**Omni$_3$**) For all provable B that do not depend on an A_i, $\Box_n B$ for $n > 0$.

By the line of reasoning employed in the critique of the Joint Disprovability Model, (Omni$_3$) then entails that any agent that can perform just 1 step in $\mathcal{S}$ is logically omniscient with respect to any logical truth B that does not depend on an A_i.

Strictly, we do not have full blown logical omniscience, and strictly we do not get counterexamples to the Carnap-style analyses of $\Diamond_n$ and $\Box_n$. In both cases, there are logically true sentences that may be false at scenarios in W_n, for $n > 0$. The relevant examples are found by looking at sentences that involve at least one of the atomics A_i. But clearly (Omni$_3$) is a counterintuitive consequence, and it fares badly with the intended role that $\Box_n$ is supposed to play in an analysis of what it means to *easily* establish a sentence a priori. And in particular, since (Omni$_3$) holds, we cannot use Intermediate Models $\mathcal{B}$ to make progress on the content and rationality problems.

For the content desideratum, we want to use epistemic n-intensions defined over scenarios in W_n to model the contents of the epistemic states of moderately ideal agents. We cannot use Intermediate Models $\mathcal{B}$ to satisfy the content desideratum. To see this, let W_n have level-m, for some finite m, and let the relevant set of inconsistencies all involve one of the atomic sentences $A_1, A_2, \ldots, A_m$. First, consider any n-provable $(A \leftrightarrow B)$, where contingent A involves at least one of the atomics A_i. To capture the basic inferential relations between $(A \leftrightarrow B)$ and A and B, we want the n-intensions of A and B to coincide in truth-value when the n-intension of $(A \leftrightarrow B)$ is necessary. But they do not in Intermediate Models $\mathcal{B}$: Though the n-intension of $(A \leftrightarrow B)$ is true at all scenarios in W_n, the values of the n-intensions of A and B can come arbitrarily apart at many scenarios in W_n. So we cannot use the n-intensions that we can define in Intermediate Models $\mathcal{B}$ to reflect the basic inferential relations that may hold between sentences involving a given A_i.

Second, consider any sentence B that does not involve any of the atomics A_i. By (Omni$_3$), if B is provable in $\mathcal{S}$, and if B does not depend on an A_i, then $\Box_n B$ for $n > 0$. By (C-$\Box_n$), then B is true at all $w \in W_n$, and derivatively the n-intension of B is necessary for any $n > 0$. But if we assume that all moderately ideal agents can at least perform 1 step in $\mathcal{S}$, this is clearly the wrong result: Moderately ideal agents may well fail to accept some arbitrarily complex logical truth B that does not depend on an A_i. Further, consider any two contingent sentences B and C that do not involve an A_i, and assume C is logically equivalent to B. If the n-intension of B is true at some $w \in W_n$, then the n-intension of C is also true at w. But again, this is the wrong result: If the logical equivalence of B and C is highly non-trivial and computationally demanding to establish, then a moderately ideal agent may well fail to accept C, even though she accepts B. Accordingly, we cannot use the n-intensions that we can define in Intermediate Models $\mathcal{B}$ to represent the contents of the epistemic states of moderately ideal agents. So we cannot use Intermediate Models $\mathcal{B}$ to satisfy the content desideratum.

For the rationality desideratum, we want to use scenarios in W_n to analyze a non-trivial notion of epistemic possibility that captures which sentences should remain epistemically possible for minimally rational agents. We cannot use Intermediate Models $\mathcal{B}$ to satisfy the rationality desideratum. To see this, let again W_n have level-m, for some finite n and m greater than 0, and let the relevant set of inconsistencies all involve one of the atomic sentences $A_1, A_2, \ldots, A_m$. First, consider any sentence A that involves at least one of the atomics A_i. We want to use scenarios in W_n to say that if $(A \wedge B)$ is epistemically necessary for a minimally rational agent a, then A should also be epistemically necessary for a. But we cannot: For any W_n, though $(A \wedge B)$ is true at each $w \in W_n$ that remains epistemically possible for a, there is no guarantee that A is also true at w. When A involves at least one of the atomics A_i, the construction allows that scenarios w such that $\{(A \wedge B), \neg A\} \subset w$ may remain epistemically possible for a. Second, for any sentence B that does not depend on any A_i, we cannot use scenarios in W_n to infer claims about which sentences should remain epistemically possible for a. By (Omni$_3$), this is immediately clear: Any minimally rational agent is wrongly characterized as logically omniscient with respect to any logical truth B that does not depend on any A_i. So we cannot use Intermediate Models $\mathcal{B}$ to satisfy the rationality desideratum.

So with respect to the content and rationality desiderata, Intermediate Models $\mathcal{B}$ fare no better than Intermediate Models $\mathcal{A}$. As we have seen, either scenarios in Intermediate Models $\mathcal{B}$ remain too unconstrained to play the role of non-trivially impossible scenarios, or logical omniscience sneaks back in. In fact, there is reason to say that Intermediate Models $\mathcal{B}$ fare worse than Intermediate Models $\mathcal{A}$ because of (Omni$_3$). That is, whereas worries about logical omniscience are bound to arise for a wide class of agents and sentences in Intermediate Models $\mathcal{B}$, these worries could at least only arise under fairly specific conditions in Intermediate Models $\mathcal{A}$. But in any case, neither Intermediate Models $\mathcal{A}$ nor Intermediate Models $\mathcal{B}$ can play the role of non-trivial epistemic space.

Though I tied the discussion of Intermediate Models to the Single Disprovability Model, the negative points apply more generally. If we want to say that some, but not all trivially impossible scenarios can survive in epistemic space, we have two options. First, we can aim to restrict the number of blatant inconsistencies that each inconsistent scenario in non-ideal epistemic space can contain. Second, we can aim to secure that some kinds of blatant inconsistencies are never contained by any scenario in non-ideal epistemic space. As exemplified by Intermediate Models $\mathcal{A}$ and $\mathcal{B}$ respectively, both options are of little use for making progress on the content and rationality problems. So when scenarios are maximal, I conclude that there is no successful construction of a non-trivial epistemic space in which some, but not all trivially impossible scenarios may survive.

5.3 Partial Models

I have investigated the first option that remains open if we accept the first horn of the dilemma. I now turn to the second option and investigate models of non-ideal epistemic space in which scenarios fail to obey (Maximality):

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

Call models of epistemic space in which scenarios fail to obey (Maximality) *Partial Models*, and call scenarios that fail to be maximal *partial scenarios*.

If we admit partial scenarios in non-ideal epistemic space, we can allow that sentences may be indeterminate in truth-value at scenarios. For instance, though A and $(A \rightarrow B)$ are true at some partial scenario w, both B and $\neg B$ may remain indeterminate at w. And if we admit partial scenarios in non-ideal epistemic space, results like (Omni) and (Incon) need not create any immediate problems, since these results rely on scenarios being maximal.[15] Of course, (Omni) and (Incon) do not directly threaten Extreme Epistemic Space and the Single Disprovability Model either. These results only start to bite when we want to do better than Extreme Epistemic Space and the Single Disprovability Model, and in particular when we want to make progress on the content and rationality problems.

Obviously, we do not make progress on the content and rationality problems *merely* by including partial scenarios in Extreme Epistemic Space or the

[15] (Omni) is the result from chapter 4, which says that for all sentences B and $n > 1$, if B is provable in $\mathcal{S}$, then $\boxtimes_n B$, for some operator $\boxtimes_n$ that satisfies the Carnap-style analysis.

Single Disprovability Model. We need something more from Partial Models. In the previous chapter, I investigated non-maximal worlds in the broad setting of Jago's model of non-ideal epistemic space. In this section, I investigate partial scenarios in the broad setting of the Single Disprovability Model. More specifically, since there is good motivation for saying that trivially impossible scenarios remain epistemically impossible for moderately ideal agents, I will investigate whether we can use Partial Models to eliminate all such trivially impossible scenarios from epistemic space while steering clear of logical omniscience. If successful, we can then investigate whether we can use Partial Models to satisfy the content and rationality desiderata.

To develop a *Partial Model*, I first identify a scenario w with an arbitrary set of sentences in $\mathcal{L}$—that is, with an arbitrary set of possible sentence types in English*. Second, I define what it means for a sentence A to be true or false at a scenario as follows:

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

(**Falsity***) A sentence A is false at scenario w iff $\neg A \in w$.

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 or verifies $\neg A$. If A is neither true nor false at w, I will say that A is *indeterminate* at w.[16]

Third, I add:

> (**Joint Disprovability**) A set of sentences Γ can be jointly disproved within n steps in $\mathcal{S}$ iff a contradiction can be derived from Γ within n steps in $\mathcal{S}$.

As in the Joint Disprovability Model, we can then define the following notion of consistency:

[16] As in Extreme Epistemic Space, we can also easily invoke the definition of what it means for two scenarios to be equivalent and derive (Parsimony).

> (n-**Consistency**) A set of sentences Γ is n-consistent with respect to $\mathcal{S}$ iff Γ cannot be jointly disproved within n steps in $\mathcal{S}$; otherwise Γ is n-inconsistent with respect to $\mathcal{S}$.

So a scenario w is n-consistent just in case we cannot derive a contradiction from w within n steps in $\mathcal{S}$.

Fourth, I utilize the idea behind the construction of W_n from the Single Disprovability Model and define (n-Saturation), where PR_n is the set of n-provable sentences in $\mathcal{S}$:

> (n-**Saturation**) A set of sentences Γ is n-saturated iff $PR_n \subseteq \Gamma$; otherwise Γ is n-unsaturated.

So a scenario w is n-saturated just in case w contains all sentences that are provable within n steps in $\mathcal{S}$. If $n = 0$, then by stipulation $PR_n = \emptyset$, in which case all scenarios are trivially 0-saturated.

Let V_P be the class of scenarios that satisfy the principles above. We can then define the following spheres of scenarios:

$$\begin{aligned} V_0 &= \{w \in V_P |\ w \text{ is 0-consistent and 0-saturated}\}. \\ &\ \vdots \\ V_n &= \{w \in V_P |\ w \text{ is } n\text{-consistent and } n\text{-saturated}\}. \\ &\ \vdots \\ V_\infty &= \{w \in V_P |\ w \text{ is } \infty\text{-consistent and } \infty\text{-saturated}\}. \end{aligned}$$

Since I stipulate that no sentence A is provable in 0 steps in $\mathcal{S}$, V_P is identical to the class of scenarios V_0. V_n corresponds to the class of scenarios in V_0 that cannot be jointly disproved within n steps in $\mathcal{S}$ and that contain all n-provable sentences in $\mathcal{S}$. V_∞ corresponds to the class of scenarios in V_0 that cannot be jointly disproved in any number of steps in $\mathcal{S}$ and that contain all provable sentences in $\mathcal{S}$.

By (Incon) and (Joint Disprovability), all maximal, logically inconsistent scenarios are eliminated from V_n for small n. Hence all trivially impossible scenarios are quickly eliminated from epistemic space. Yet, since there are plenty of partial scenarios in V_n, we can avoid logical omniscience. For instance, take a sentence B that is provable in 100 steps in $\mathcal{S}$. Consider then two scenarios w_1 and w_2 in V_{20}, where A, C, and D are neither provable nor disprovable within 20 steps in $\mathcal{S}$:

$$w_1 = PR_{20} \cup \{C, (C \rightarrow D), \neg B\}.$$

$$w_2 = PR_{20} \cup \{C, (C \rightarrow D), \neg A\}.$$

Neither w_1 nor w_2 is jointly disprovable within 20 steps in $\mathcal{S}$, and both scenarios contain all 20-provable sentences. But since B is false at w_1 and indeterminate at w_2, we can avoid saying that all $w \in V_{20}$ verify B. Derivatively, we can steer clear of logical omniscience. Of course, since all *maximal* scenarios that verify $\neg B$ are eliminated from V_{20} by (Joint Disprovability), all maximal scenarios in V_{20} verify B.[17] But when partial scenarios are allowed in V_{20}, this creates no immediate problems.

The (n-Saturation) principle is there to help us capture the basic picture that guides constructions of non-trivial epistemic space. For that picture, the Carnap-style analyses (C-$\Diamond_n$) and (C-$\Box_n$) are essential. Though the (C-$\Diamond_n$) analysis is troublesome when partial scenarios are admitted in V_n, we can use (n-Saturation) to "fill up" scenarios in V_n with enough sentences to establish (C-$\Box_n$):

(C-$\Box_n$) $\Box_n A$ iff A is true at all scenarios w in V_n.

[17] As always, I here assume that $\mathcal{S}$ is a reasonable system, in which all maximal sets of sentences that contain a LNC-, MP-, or NC-inconsistency can be ruled out within 20 steps.

Without (n-Saturation), we would not be able to prove (C-$\Box_n$) because scenarios like $w = \{A, B\}$, for compatible sentences A and B, would never be eliminated from V_n.[18] But with (n-Saturation), the proof of (C-$\Box_n$) proceeds exactly as the proof of (C-$\Box_n$) in the Single Disprovability Model.[19]

Yet, it is also easily seen why we *cannot* use the construction of V_n to prove (C-$\Diamond_n$):

> **(C-**$\Diamond_n$) $\Diamond_n A$ iff A is true at some scenario w in V_n.

For (C-$\Diamond_n$) left to right, assume $\Diamond_n A$. Since $\Diamond_n A$ is defined as $\neg\Box_n\neg A$, then $\neg A$ is not provable in n steps in $\mathcal{S}$. So $\neg A \notin PR_n$. By construction, there is hence a $w \in V_n$ such that $\neg A \notin w$. But since scenarios in V_n can be partial and A indeterminate at w, we cannot make the inference from $\neg A \notin w$ to $A \in w$. Given the definition of '$\Diamond_n$' as $\neg\Box_n\neg$, we hence cannot use V_n to establish (C-$\Diamond_n$); similar problems arise for the right to left direction in (C-$\Diamond_n$).[20]

To establish (C-$\Diamond_n$) in Partial Models, there seems to be two options. First, we aim to change the scenario-involving analysis of $\Diamond_n$. Second, we aim to separate the notion of disprovability from the notion of provability and take both as primitive notions.

On the first option, consider for instance the following analysis:

[18] By aiming to eliminate scenarios like $w = \{A, B\}$ by use of non-classical, many-valued inference rules, we saw in section 4.3.1, chapter 4, that all scenarios that fail to be logically consistent are eliminated in few steps. To avoid such results while simultaneously ensuring (C-$\Box_n$), we "fill up" scenarios manually instead.

[19] That is:

> For (C-$\Box_n$) left to right, assume $\Box_n A$. Then A is provable in n steps in $\mathcal{S}$. So $A \in PR_n$. By construction of V_n, then $A \in w$ for all $w \in V_n$. For (C-$\Box_n$) right to left, assume $A \in w$ for all $w \in V_n$. For reductio, assume $A \notin PR_n$. Then, by construction of V_n, there is a $w \in V_n$ such that $A \notin w$. By assumption, there is no such w, so $A \in PR_n$. Then A is provable in n steps in $\mathcal{S}$, and so $\Box_n A$. So (C-$\Box_n$) holds. □

[20] Notice, however, that there are no problems with the proofs (Epi-Pos) and (Epi-Nec). By taking both (occurrent) acceptance and (occurrent) rejection as primitive notions, we can prove (Epi-Pos) and (Epi-Nec) in Partial Models exactly as we proved them in Extreme Epistemic Space.

(**C-**$\Diamond'_n$) $\Diamond_n A$ iff there is a $w \in V_n$ such that $\neg A \notin w$.

We can then immediately prove (C-$\Diamond'_n$):

> For (C-$\Diamond'_n$) left to right, assume $\Diamond_n A$. Then $\neg\Box_n\neg A$. Then $\neg A$ is not provable in n steps in $\mathcal{S}$, and hence $\neg A \notin PR_n$. So by construction of V_n, there is some $w \in V_n$ such that $\neg A \notin w$. From right to left, assume there is a $w \in V_n$ such that $\neg A \notin w$. Then not for all $w \in V_n$, $\neg A \in w$. By construction of V_n, then $\neg A \notin PR_n$. Then $\neg A$ is not provable in n steps in $\mathcal{S}$. Hence $\neg\Box_n\neg A$, and so $\Diamond_n A$. So (C-$\Diamond'_n$) holds. □

By employing (C-$\Diamond'_n$), we effectively give up the basic Carnap-style analysis of $\Diamond_n$, which requires that A is *true* at some $w \in V_n$ when $\Diamond_n A$ holds. With (C-$\Diamond'_n$), the fact that A is *not false* at any $w \in V_n$ is enough for $\Diamond_n A$ to hold. But a "possibility operator, it could be argued, would not be as flabby as this."[21] Yet (C-$\Diamond'_n$) is an option that adherents of Partial Models may adopt to ensure a scenario-involving analysis of $\Diamond_n$.

On the second option, we take 'disprovability in n steps in $\mathcal{S}$' as primitive. For the purpose of establishing (C-$\Diamond_n$), one needs a distinction between 'provability' and 'disprovability' that mirrors the distinction between acceptance and rejection, where both (occurrent) acceptance and (occurrent) rejection are taken as primitive.[22] In particular, one needs a notion of n-disprovability such that if A is disprovable in n steps in $\mathcal{S}$, then $A \notin w$ for any $w \in V_n$. Given a notion of n-disprovability, one can then introduce a corresponding operator

[21] Schotch et al. (1978): p. 66, with respect to an analysis of $\Diamond$ similar to (C-$\Diamond'_n$).

[22] There might be some hints on how to separate 'provable $\neg A$' from 'disprovable A' in the recent literature on "rejective negation". Roughly, one might aim to ground a distinction between 'provable $\neg A$' and 'disprovable A' by arguing for a relevant distinction between 'accepting $\neg A$' and 'rejecting A'. For discussions on the relations between the notions of acceptance, rejection, and classical and non-classical negation, see Humberstone (2000), Priest (2006a), Priest (2006b), Restall (2005a), Rumfitt (2000), and Simley (1996).

that means 'disprovability in n steps in $\mathcal{S}$' and derivatively define $\Diamond_n$ as the negation of this operator.

Assuming that a plausible distinction between provability and disprovability has been drawn, we can then let DR_n be the set of sentences that are disprovable within n steps in $\mathcal{S}$ and define:

> (n-**Irrefutability**) A set of sentences Γ is n-irrefutable iff $DR_n \not\subseteq \Gamma$; otherwise Γ is n-refutable.

We can then complicate the construction of V_n as follows:

> $V'_n = \{w \in V'_0 |\ w$ is n-consistent, n-saturated, and n-irrefutable$\}$.

Then V'_n is the class of scenarios, which cannot be jointly disproved within n steps, which contain all n-provable sentences, and which contain no n-disprovable sentences. We have now built everything into V'_n that we need to establish (C-$\Diamond_n$):[23]

> (**C-**$\Diamond_n$) $\Diamond_n A$ iff A is true at some scenario w in V'_n.

The proof of (C-$\Diamond_n$) is immediate:

> For (C-$\Diamond_n$) left to right, assume $\Diamond_n A$. Then A is not disprovable in n steps in $\mathcal{S}$. So $A \notin DR_n$. By construction of V'_n, then it is not the case that $A \notin w$ for each $w \in V'_n$. So there is a $w \in V'_n$ such that $A \in w$. For (C-$\Diamond_n$) right to left, assume $A \in w$ for some $w \in V'_n$. For reductio, assume $A \in DR_n$. By construction of V'_n, then $A \notin w$ for any $w \in V'_n$. But by assumption, $A \in w$ for some $w \in V'_n$. So $A \notin DR_n$. Hence A is not disprovable in n steps in $\mathcal{S}$, and so $\Diamond_n A$. So (C-$\Diamond_n$) holds. □

[23] The proof of the corresponding version of (C-$\Box_n$) is identical to the proof of (C-$\Box_n$) in the Single Disprovability Model, so I skip it.

If 'disprovable A' does not mean 'provable $\neg A$', we need, of course, another explanation of what disprovability amounts to. Such an explanation will be non-obvious if we hold on to propositional logic and classical negation. But rather than dwelling on the possible interpretations of 'disprovability', I will leave the V'_n construction as an option that adherents of Partial Models may adopt to ensure a scenario-involving analysis of $\Diamond_n$.

So in contrast to models of epistemic space in which scenarios are maximal, complications immediately arise in Partial Models with respect to the basic (C-$\Diamond_n$) analysis. To ensure (C-$\Diamond_n$) or something in the vicinity, we either lose the simple Carnap-style analysis of $\Diamond_n$, or we have to re-interpret the notion of disprovability. Insofar as we can isolate a diamond operator that can play the same role as $\Diamond_n$ and be interpreted in terms of existential quantification over scenarios in V_n, these complications need not carry substantial costs. But unless Partial Models allow us to make substantial progress on the content and rationality problems, they are nonetheless costs that models based on maximal scenarios do not have. If we cannot ensure that A is *true* at some $w \in V_n$ when $\Diamond_n A$ holds, we lose a core feature of the traditional world involving framework—a feature that we may otherwise retain even when non-ideal scenarios or impossible worlds are admitted in modal space. And if we cannot interpret 'A is disprovable' as '$\neg A$ is provable', we lose a core conceptual feature of classical logic.

But for the broad picture, I will assume that we have a reasonable way to establish something akin to (C-$\Diamond_n$) and instead evaluate whether we can use Partial Models to satisfy the content and rationality desiderata.

5.3.1 Problems in Partial Models

To avoid suffering the same fate as the Joint Disprovability Model, it is essential to Partial Models that scenarios like $w = PR_n \cup \{A, (A \rightarrow B)\}$ are never eliminated from V_n, where A and B are contingent sentences. I will call scenarios such as $w = PR_n \cup \{A, (A \rightarrow B)\}$, where B is indeterminate, and scenarios such as $w' = PR_n \cup \{A, (A \wedge B)\}$, where B is indeterminate, *defective scenarios*. We have principles in Partial Models that can eliminate from V_n all scenarios that contain $\{A, (A \rightarrow B), \neg B\}$, and in general all scenarios that contain a LNC-, MP- or NC-inconsistency. But to avoid the omniscience problems that we face in the Joint Disprovability Model, defective scenarios *must* survive in Partial Models. Effectively, we avoid logical omniscience by replacing trivially impossible scenarios with defective scenarios in V_n. To illustrate, consider a scenario $w \in V_4$, where A through D are contingent sentences:

$$w = \; PR_4 \cup \{A, (A \rightarrow B), (B \rightarrow C), (C \rightarrow D), \neg D\}.$$

Though w contains all 4-provable sentences, there is no standard way to eliminate w in less than 4 steps.[24] *Had* B been false at w, w could be jointly disproved in 2 steps in $\mathcal{S}$. *Had* B been true at w, w could be jointly disproved in 3 steps. But since B is indeterminate at w, we can avoid the kind of reasoning that we used to criticize the Joint Disprovability Model. Roughly, if we gave each indeterminate sentence in a defective scenario a truth-value, we would be able to collapse the spherical structure of Partial Models as we did with the Joint Disprovability Model. As we saw in the previous chapter, this also tells against allowing $\mathcal{S}$ to reason about indeterminacies. If there are rules in $\mathcal{S}$ that enable us to eliminate defective scenarios from V_n, logical omniscience will sneak right back in.

Rather than discussing the intuitive and dialectical strength of replacing trivially impossible scenarios with defective scenarios, let us evaluate Partial Models with respect to the content and rationality problems. Generally speaking-

[24] I count the application of the trivial rule that allows us to infer A in 1 step from any set Γ such that $A \in \Gamma$, but if we do not bother about this rule, subtract 1 throughout the following cases.

ing, Partial Models approach these problems by replacing cases of explicit inconsistencies with cases of indeterminacies.

For the content desideratum, we want to use non-ideal epistemic intensions defined over scenarios in V_n to represent the contents of the epistemic states of moderately ideal agents. Let the *partial n-intension* of a sentence be a function from scenarios in V_n to a truth-value. Because of (n-Saturation), we are immediately guaranteed that the partial n-intension of any n-provable sentence is necessary. But as we know, the real battle takes place when we look beyond the class of n-provable sentences.

So consider the simple inference from $(A \wedge B)$ to A, where A and B are contingent sentences. To reflect the basic inferential relations among such sentences and thoughts in the corresponding partial n-intensions, we want the partial n-intension of A to be true at each scenario $w \in V_n$, for sufficiently large n, whenever the partial n-intension of $(A \wedge B)$ is true at w. Partial Models cannot do this job: For any V_n, though $(A \wedge B)$ is true at some $w \in V_n$, there is no guarantee that A is also true at w. Since there are plenty of partial scenarios in V_n, A might be indeterminate at w though $(A \wedge B)$ is true at w. To be sure, the partial n-intension of A can never be *false* when the partial n-intension of $(A \wedge B)$ is *true*: Any scenario w that contains $\{(A \wedge B), \neg A\}$ is quickly eliminated from V_n by (Joint Disprovability). But still, to represent the cognitively trivial inference from $(A \wedge B)$ to A, we want the non-ideal intension of A to be *true* when the non-ideal intension of $(A \wedge B)$ is true. Since partial n-intensions cannot do this job, they cannot play the role of non-ideal epistemic intensions.

Or consider the basic inferential relations between $(A \leftrightarrow B)$ and A and B, where A and B are contingent sentences. Assume $(A \leftrightarrow B)$ is provable within n steps in $\mathcal{S}$ and hence true at each $w \in V_n$. Then the partial n-intension of $(A \leftrightarrow B)$ is necessary. If both A and B have a truth-value at any $w \in V_n$, then they must coincide: Any scenario w that contains $\{(A \leftrightarrow B), \neg A, B\}$ is quickly eliminated from V_n by (Joint Disprovability). Yet, since there are plenty of partial scenarios in V_n, the partial n-intension of B need never be *true* when the partial n-intension of A is true and the partial n-intension of $(A \leftrightarrow B)$ is necessary. But if the logical equivalence of A and B can be established in a sufficiently small number of steps in $\mathcal{S}$, then intuitively any moderately ideal agent is a priori certain that A and B are logically equivalent. If this moderately ideal agent also accepts A, she will also accept B because of the basic and obvious inferential relations that exist between these sentences. Derivatively, we want the non-ideal intension of B to be true when the non-ideal intension of A is true and the non-ideal intension of $(A \leftrightarrow B)$ is necessary. Since partial n-intensions cannot do this job, they cannot play the role of non-ideal epistemic intensions.

More generally, when we want to make positive claims about the structural features of the epistemic states of moderately ideal agents, Partial Models do not improve on previous models. This is not surprising. Except for containing all sentences that are provable within n steps in $\mathcal{S}$, *partial* scenarios can fail to verify any sentence that is an obvious logical consequence of other sentences that these scenarios verify. But if we want to use scenarios in Partial Models to give a world involving analysis of a non-trivial notion of hyperintensional content, this is not good enough. For that job, we want to use scenarios in non-ideal epistemic space to capture obvious or easy deductive inferences in thought. But since defective scenarios like $PR_n \cup \{A, (A \rightarrow B)\}$ and $PR_n \cup \{(A \wedge B)\}$ need never be eliminated from *any* V_n, when A and B are contingent sentences, scenarios in V_n cannot do this job. As in the Single Disprovability Model, the kinds of non-ideal epistemic intensions that we can isolate in Partial Models remain too unconstrained to capture cognitively trivial or computationally feasible inferences among sentences and thoughts. So we cannot use Partial Models to satisfy the content desideratum.

For the rationality desideratum, we want to use scenarios in V_n to analyze a non-trivial notion of epistemic possibility that captures which sentences should remain epistemically possible for minimally rational agents. Consider again two contingent sentences A and B. Suppose a minimally rational agent a accepts $(A \wedge B)$. Since a can easily infer A from $(A \wedge B)$, she rationally should accept A when she accepts $(A \wedge B)$. We capture this normative component by saying that if $(A \wedge B)$ is epistemically necessary for a, then so is A. By merely replacing trivially impossible scenarios with defective scenarios, we cannot use Partial Models to capture such minimal constraints on rational acceptance: For any V_n, though $(A \wedge B)$ is true at each $w \in V_n$ that remains epistemically possible for a, there is no guarantee that A is also true at w. Since there are plenty of partial scenarios in V_n, A might be indeterminate at w though $(A \wedge B)$ is true at w. To be sure, if $(A \wedge B)$ is epistemically necessary for a, then it can never be the case that $\neg A$ is also epistemically necessary for a. But still, we cannot use scenarios in V_n to infer that A should be epistemically necessary for a whenever $(A \wedge B)$ is. As in the Single Disprovability Model, we hence cannot use scenarios in Partial Models to infer basic claims about which sentences should remain epistemically possible for minimally rational agents. So we cannot use Partial Models to satisfy the rationality desideratum.

So although Partial Models allow us to eliminate all trivially impossible scenarios from V_n, the presence of defective scenarios in V_n still prevent us from making substantial progress on the content and rationality problems. So even if we abstract away from problems concerning the (C-$\Diamond_n$) analysis, Partial Models cannot play the role of non-trivial epistemic space.

Given this, we might wonder whether we can improve on Partial Models by utilizing the ideas from Intermediate Models. I will very briefly argue why we cannot. Let a partial scenario w be *m-indeterminate* just in case all indeterminacies in w involve at most m many different atomic sentences $A_1, A_2, \ldots, A_m$. For instance, if partial w is 1-indeterminate, then all indeterminacies in w might only involve A. When w is m-indeterminate, then for all sentences B that do not involve any A_i, either B is true at w or B is false at w.

With a view to Intermediate Models $\mathcal{A}$, let us stipulate that each partial scenario in V_n is at most m-indeterminate, for some finite n and m greater than 0. For all atomic sentences A, there is then some partial scenario $w \in V_n$ whose indeterminacies involve A. For the content desideratum, we need the following: If $(A \leftrightarrow B) \in w$ for all $w \in V_n$, and $B \in w'$ for some $w' \in V_n$, then $A \in w'$, where A is atomic.[25] But since there are plenty of scenarios in V_n whose indeterminacies involve A, we are never guaranteed that $A \in w$ just because $\{(A \leftrightarrow B), B\} \subset w$. So scenarios in V_n still lack the required structure for defining partial n-intensions that can play the role of non-ideal epistemic intensions. For the rationality desideratum, we need the following: If $(A \wedge B) \in w$, for some $w \in V_n$, then $A \in w$, where A is atomic. But again, since there are plenty of partial scenarios in V_n whose indeterminacies involve A, we are never guaranteed that $A \in w$ when $(A \wedge B) \in w$ for each w that remains epistemically possible for a minimally rational agent. So we still cannot use scenarios in V_n to infer basic claims about which sentences should remain epistemically necessary for minimally rational agents.

With a view to Intermediate Models $\mathcal{B}$, suppose each sphere V_n is m-indeterminate, for some finite n and m greater than 0. For each partial scenario w in V_n, w is then at most indeterminate with respect to sentences that involve the atomics $A_1, A_2, \ldots, A_m$. When V_n is m-indeterminate, for some sufficiently small n, then each $w \in V_n$ is maximal and consistent with respect to any B

[25] For the general critique, it does not matter that A is atomic, but it facilities the discussion.

that does not involve an A_i.[26] If we focus on the content desideratum, this version of a Partial Model cannot give us what we want. First, for some small $n > 0$, the partial n-intension of any provable B that does not depend on an A_i is necessary. But then any agent that can perform just this small number of steps in $\mathcal{S}$ is logically omniscient with respect to any such B, which is the wrong result. Second, consider any n-provable $(A \leftrightarrow B)$, where A involves an A_i. If $(A \leftrightarrow B) \in w$ for all $w \in V_n$, and $B \in w'$ for some $w' \in V_n$, we need $A \in w'$. But when V_n is m-indeterminate and A involves an A_i, we are never guaranteed that $A \in w$ just because $\{(A \leftrightarrow B), B\} \subset w$. So we still cannot use scenarios in V_n to define partial n-intensions that can play the role of non-ideal epistemic intensions.

If we focus on the rationality desideratum, the current version of a Partial Model does not give us what we want either. First, consider any B that does not involve an A_i. For this class of sentences, we cannot use scenarios in V_n to make claims about which sentences should remain epistemically possible for minimally rational agents. As above, this is immediately clear since such agents are wrongly characterized as logically omniscient with respect to all logical truths that do not depend on an A_i. Second, consider $(A \wedge B)$, where A and B are contingent, and where A involves an A_i. If $(A \wedge B) \in w$, we want $A \in w$, for all $w \in V_n$, where V_n is m-indeterminate. But since there are partial scenarios in V_n whose indeterminacies involve an A_i, there is no guarantee that $A \in w$ whenever $(A \wedge B) \in w$ for each scenario w that remains epistemically possible for a minimally rational agent. So we still cannot use scenarios in V_n to infer basic claims about which sentences should remain epistemically possible and necessary for minimally rational agents.

[26] The sufficiently small value n is the value for which all instances of a LNC-, MP-, and NC-inconsistency can be disproved in $\mathcal{S}$.

So even in combination with some of the techniques used in Intermediate Models, we cannot use Partial Models to satisfy the content and rationality desiderata. To avoid logical omniscience in Partial Models, defective scenarios must survive in epistemic space. But as we have seen, the presence of such defective scenarios means that Partial Models cannot play the role of non-trivial epistemic space. So I conclude that there is no successful construction of non-trivial epistemic space when scenarios may be partial.

We want a non-trivial epistemic space that can help us capture basic logical inferences among thoughts and sentences, while simultaneously help us steer clear of logical omniscience. Like the simple Single Disprovability Model, Partial Models allow us to steer clear of logical omniscience. But like the Single Disprovability Model, Partial Models do not enable us to capture basic logical inferences among thoughts and sentences. For a construction of a non-ideal, yet *non-trivial* epistemic space, this does not suffice.

5.4 Summary

When scenarios are maximal, I conclude by (Incon) that there is no construction of a non-ideal epistemic space that contains no trivially impossible scenarios. In light of the dilemma, we must hence admit trivially impossible scenarios in epistemic space to avoid logical omniscience. For current purposes, this leaves us with two options: Intermediate Models and Partial Models.

As we saw from the discussion of Intermediate Models, epistemic spaces that contain some, but not all trivially impossible scenarios cannot play the role of non-trivial epistemic space. In Intermediate Models $\mathcal{A}$, scenarios remain too unconstrained to play the role of non-trivially impossible scenarios. In Intermediate Models $\mathcal{B}$, either scenarios remain too unconstrained to play the role of non-trivially impossible scenarios, or logical omniscience sneaks back in. So when scenarios are maximal, I have argued that there is no successful construction of a non-ideal, yet non-trivial epistemic space.

As we saw from the discussion of Partial Models, defective scenarios must survive in epistemic space to avoid logical omniscience. But if so, partial scenarios remain too unconstrained to play the role of non-trivially impossible scenarios. So when scenarios may be partial, I have argued that there is no successful construction of a non-ideal, yet non-trivial epistemic space.

Since Intermediate Models and Partial Models are the only options left open by the horns of the dilemma, I have hence given reasons for thinking that there is no successful construction of a non-trivial epistemic space. And at least I hope to have shown that successful constructions of non-trivial epistemic spaces are hard, if not impossible to find.

In the final chapter, I conclude and discuss some of the ramifications of these results.

Chapter 6

Conclusion

Throughout the previous chapters, I have argued for the following claims:

> ($\mathbf{R}_1$) When scenarios are maximal, there is no construction of an epistemic space W such that:
>
> (i) W contains only possible and non-trivially impossible scenarios, and such that
>
> (ii) W allows us to model agents that are not logically omniscient.
>
> ($\mathbf{R}_2$) When scenarios are maximal, there is no construction of an epistemic space W such that:
>
> (i) W contains some but not all trivially impossible scenarios, and such that
>
> (ii) W allows us to satisfy the content and rationality desiderata.
>
> ($\mathbf{R}_3$) When scenarios may fail to be maximal, there is no construction of an epistemic space W such that W allows us to satisfy the content and rationality desiderata.

(R$_1$) follows from (Incon), and it captures the core of the dilemma from chapter 5 to the effect that W must contain trivially impossible scenarios if we want to avoid logical omniscience. As we saw, the dilemma is vicious for the general picture that motivates models of non-trivial epistemic space. If W contains trivially impossible scenarios, then W cannot play the role of non-trivial epistemic space—as exemplified by the discussion of the Single Disprovability Model in chapter 4. If W does not contain any trivially impossible scenarios, then W only allows us to model logically omniscient agents—as exemplified by the discussion of the Joint Disprovability Model in chapter 4.

(R$_2$) gains support from the discussion of Intermediate Models $\mathcal{A}$ and $\mathcal{B}$ in chapter 5. Whereas scenarios in Intermediate Models $\mathcal{A}$ are too unconstrained to play the role of non-trivially impossible scenarios, scenarios in Intermediate Models $\mathcal{B}$ either are too unconstrained to play the role of non-trivially impossible scenarios, or they are suitable only for modeling logically omniscient agents. In either case, we cannot use the models to satisfy the content and rationality desiderata.

(R$_3$) gains support from the discussion of Jago's model in chapter 4 and from the discussion of Partial Models in chapter 5. If we want to satisfy the content and rationality desiderata, we first of all need to avoid logical omniscience. But as we saw, agents characterized by Jago's model—or its reasonable precisifications—turn out to be logically omniscient. To avoid logical omniscience in models of epistemic space based on partial scenarios, we know that defective scenarios can never be eliminated from epistemic space.[1] But if so, scenarios in Partial Models are too unconstrained to play the role of non-trivially impossible scenarios. So in both cases, we cannot use the models to satisfy the content and rationality desiderata.

[1] Defective scenarios, we remember, are scenarios such as $w = PR_n \cup \{A, (A \rightarrow B)$, where B is indeterminate, and scenarios such as $w' = PR_n \cup \{A, (A \wedge B)\}$, where B is indeterminate, and where PR_n is the set of n-provable sentences in some system $\mathcal{S}$.

So if the criterion of success is measured in terms of the content and rationality desiderata, (R_1) through (R_3) constitute an argument that there are no successful constructions of non-trivial epistemic space. In turn, we have an argument that shows that the general idea behind models of non-trivial epistemic space from chapter 3 cannot be made precise. Generally, this idea requires that we can use scenarios in epistemic space to capture what can be easily established and easily ruled out a priori. But as we have seen, models of epistemic space either fail to capture what can be easily established a priori—confer with the discussions of Extreme Epistemic Space, the Single Disprovability Model, Intermediate Models, and Partial Models—or they fail to distinguish between what can and cannot be easily established a priori—confer with the discussions of the Joint Disprovability Model and Jago's model.

As a test case, I have interpreted the notion of easily establish a priori in terms of the notion of provability in n steps. But the results above are independent of this particular test case interpretation. By (G-Incon) we know that all maximal, logically inconsistent scenarios contain an instance of a LNC-, MP-, or NC-inconsistency. And intuitively, it is very plausible to hold that LNC-, MP-, and NC-inconsistencies are the kinds of inconsistencies that can be easily ruled out by any minimally logically competent agent. If so, we reach our general conclusions without mentioning the particular n-provability interpretation. That is, insofar as moderately ideal agents can easily rule out all scenarios that verify an instance of a LNC-, MP-, or NC-inconsistency, they are wrongly characterized as logically omniscient in the corresponding world involving framework. To avoid logical omniscience, we are then left with either Intermediate Models or Partial Models and the problems that these models have. As such, the main results of this project stem from the general world involving framework rather than from any particular interpretation of the notions of easily establish and easily rule out a priori.

The results above hold for the broad class of a priori inferences and truths that involve logic. By (Incon) we know that all logically inconsistent scenarios are trivially impossible. Though this result challenges Hintikka's characterization of impossible worlds as "worlds so subtly inconsistent that the inconsistency could not be expected to be known (perceived) by an everyday logician, however competent", it does not establish that *all* kinds of a priori inconsistent scenarios are trivially impossible in any interesting sense.[2] For instance,

> [n]onideal epistemic spaces may also be useful in analyzing various specific domains, such as the moral domain. We may think that the connection between the nonmoral and the moral is ultimately a priori, or we may think that moral beliefs are ultimately not truth-evaluable, but as long as the connection and the non-truth-evaluability is not obvious, there will be an interesting hypothesis space to investigate.[3]

If such a priori connections between the non-moral and the moral exist, and if such connections can be more or less obvious, we might be able to set up an epistemic space that allows us, say, to distinguish blatantly from subtly morally impossible scenarios. The current results do not necessarily affect attempts to set up an appropriate non-trivial epistemic space for such purposes.

Yet for the broad class of a priori inferences and truths that involve logic, the results (R_1) through (R_3) support the general conclusion that successful constructions of non-trivial epistemic spaces are hard, if not impossible to find. Roughly, a non-trivial epistemic space is an epistemic space that is located somewhere between Ideal Epistemic Space and Extreme Epistemic Space. But as we have seen, whereas it is rather simple to develop epistemic

[2] Hintikka (1975): p. 478. In Hintikka's own game-theoretical approach, the degree to which contradictions manifest themselves in impossible worlds is measured by the number of quantifiers that they are nested within. Yet, all valid sentences of propositional logic are still verified by each of Hintikka's impossible worlds; see also Rantala (1975) theorem 1, p. 466, on whose urn models Hintikka's approach relies. So for the class of propositional truths, Hintikka's approach is of no help.

[3] Chalmers (forthcoming): p. 51.

spaces that can avoid the omniscience worries in Ideal Epistemic Space, it remains a formidable, if not impossible challenge to develop a non-ideal epistemic space that can also avoid the explosive “anything goes” character of Extreme Epistemic Space. As exemplified by the content and rationality desiderata, a non-ideal epistemic space where not “anything goes” should contain scenarios that allow us to capture basic, but non-trivial inferential and structural relations among sentences and thoughts. But I have argued that no model of *non-ideal* epistemic space generally allows us to do this. For instance, none of the models of non-ideal epistemic space allow us to capture even the basic inference from $(A \wedge B)$ to A in general. Alternatively, as we have seen, the constructions of epistemic space that allow us to capture basic, but non-trivial inferential and structural relations among sentences and thoughts also commit us to modeling agents that are logically omniscient. Hence I have argued that we must accept the Scylla of “anything goes” if we want to avoid the Charybdis of logical omniscience.

If this is correct, there is little hope of developing a satisfying modal space that we can use to model the broad class of ordinary reasoners that are logically competent, but not logically omniscient. Plausibly, logical competence involves the capacity to derive and accept obvious consequences of an already accepted set of sentences or thoughts. For instance, competent logical reasoners are intuitively always disposed to derive and accept B when they accept A and $(A \rightarrow B)$, though they are not generally disposed to derive and accept any old C that can be deduced from what they already accept by repeated applications of modus ponens. To model such features of bounded but non-trivial logical reasoning in a broadly world involving framework, we want a modal space W that allows us to establish claims of the form: For any A and B such that B is an obvious logical consequence of A, if A obtains at $w \in W$, then B obtains at w—or alternatively, if A obtains at w, but B fails to obtain at

w, then $w \notin W$. This requires that non-ideal scenarios or impossible worlds obey substantive constraints. But as we have seen from the discussions of the content and rationality desiderata, either these constraints are too weak to capture obvious logical inferences among sentences and thoughts, or if they are not, they entail that all logical inferences are on a par. Either way, we have reason to hold that non-ideal, yet logically competent agents are not modeled well in a broadly impossible world involving framework.

These conclusions do not necessarily affect the many alternatives that exist to an impossible world involving model of non-ideal belief and knowledge. In the introduction I mentioned a few of these alternatives, but here I will briefly consider a broad class of *syntactical* or *sentential* models of belief many of which are explicitly designed to model agents that are not logically omniscient, but nevertheless logically competent.[4] Generally, such models do not analyze beliefs as truth in all possibilities of some sort, but rather aim to represent beliefs directly by a set of sentences. Crudely put, $\mathcal{B}_a A$ holds true just in case agent a has A in its belief set S_a, where $\mathcal{B}$ is the belief operator. So if an agent a has exactly 10 beliefs, sentential models represent her epistemic state by a corresponding belief set S_a that contains 10 sentences (in some language) explicitly representing her 10 beliefs. Depending on the job at hand, different conditions can be imposed on the sentences in the belief set. For instance, we can stipulate that S_a never contains contradictory pair of sentences or that S_a obeys certain closure conditions. To illustrate, I will briefly consider three versions of a sentential model.[5]

In Konolige's *deduction model of belief*, a set of inference rules is associated with each agent a and the belief set S_a is closed under these inference

[4] See, among many others, Cherniak (1986), Eberle (1974), Jago (2009b), Restall (2005b), and Wassermann (1999).

[5] For extensive presentation and discussion of various sentential models of beliefs, see in particular Jago (2006): chapters 4 - 7, but also Fagin et al. (1995): chapter 9.

rules.[6] If an agent only has access to an incomplete set of inference rules, the agent's belief set need not be perfectly consistent nor contain all logical truths. Derivatively, the idea is, we can model non-ideal agents that do not believe all logical truths nor all logical consequences of what they already believe, but that nevertheless have the capacities for logical reasoning.

In Duc's model, a particular interpretation of dynamic logic is used to model logical reasoning.[7] Let R be a set of inference rules, and let $[R]$ and $<R>$ be the standard dynamic modalities with the intended readings "always after using rule R" and "sometimes after using rule R". In this framework, we can explicitly model non-ideal agents that have the capacities to use inference rules to generate new beliefs. For example, if K_i is the (agent indexed) knowledge operator, and if $<MP_i>$means that agent i can use modus ponens to infer B from A and $(A \rightarrow B)$, then

> [...] the idea that the agent i accepts modus ponens can be formalized by the axiom: $K_iA \wedge K_i(A \rightarrow B) \rightarrow <MP_i>K_iB$. This axiom says no more than if agent i knows A and she also knows that A implies B, then after a suitable inference step she will know B.[8]

Since there is no requirement in Duc's model that agents actually perform the kinds of inferences that they have the capacity to perform, we can avoid logical omniscience but nevertheless model agents that can engage in competent logical reasoning.

As they stand, Konolige's and Duc's models do not capture resource-bounded reasoning. But Jago presents a sentential model that does.[9] In Jago's model, we model reasoning in terms of transitions between belief states. When

[6] See Konolige (1986).

[7] See Duc (1995) and Duc (1997).

[8] Duc (1997): p. 638.

[9] See Jago (2006): chapters 5-7 and Jago (2009b). Notice, however, that it seems possible to generalize Duc's model to accommodate bounded reasoning—for instance by including a dynamic 'next step' operator in the formalism.

an agent a is in belief state S_a and can employ one of its inference rules to infer a new belief B from S_a, this is modeled by a transition from S_a to a new belief state S'_a that only differs from S_a in also containing B. Bounded reasoning is then modeled in terms of limited transitions between belief states. For instance, if a can perform n steps using its inference rules, there is a chain of n many transitions that leads from a's initial belief state S_a to a's final belief state S_a^n. Intuitively, each transition represents a step of reasoning that adds a newly inferred belief to the previous stock of beliefs, and S_a^n contains every belief that a within its resource bound can infer from its initial belief state S_a through a particular chain of reasoning. Since there are no requirements that agents can perform arbitrarily many steps using their inference rules, we can hence avoid logical omniscience but nevertheless model logically competent, resource-bounded agents.

Sentential models like those above can be given a fairly interesting relational model theory that is familiar from modal logic.[10] Here I will not dwell on the details, but rather stress two crucial aspects in which the models differ from the possible world models that I have investigated in this project. First, while possible world models analyze beliefs in terms of quantification over possibilities, sentential models represent beliefs directly in terms of membership in corresponding belief sets. Second, while possible world models analyze belief aggregation in terms of elimination of possibilities, sentential models analyze belief aggregation in terms of expanding belief sets. Since all my results presuppose these standard features of the general possible world framework, my results need not affect sentential models. So if one is antecedently sympathetic towards sentential models of belief and knowledge, one might take the conclusions of this project to indirectly motivate these models.

[10] For all the details, see Jago (2009b).

Bibliography

Adams, R. (1974): 'Theories of Actuality', in *Nous*, 8, pp. 211-231.

Armour-Garb, B., Beall, J. C., Priest, G. (eds.) (2004): *The Law of Non-Contradiction—New Philosophical Essays*, Oxford, Oxford University Press.

Bar-Hillel, Y. (1964): *Language and Information*, London, Addison-Wesley.

Bar-Hillel, Y. & Carnap, R. (1953): 'An Outline of a Theory of Semantic Information', in Bar-Hillel (1964): pp. 221-274.

Barwise, J. (1997): 'Information and Impossibilities', in *Notre Dame Journal of Formal Logic*, 38, pp. 488-515.

Beall, J. C. & Restall, G. (2000): 'Logical Pluralism', in *Australasian Journal of Philosophy*, 78, pp. 475-493.

Boghossian, P. (2003): 'The Normativity of Content', in *Philosophical Issues*, 13, pp. 31-45.

Boolos, G. (1975): 'On Second-Order Logic', in *Journal of Philosophy*, 72, pp. 509-527.

Boolos, G. (1979): *The Unprovability of Consistency: An Essay in Modal Logic*, Cambridge, Cambridge University Press.

Boolos, G. (1993): *The Logic of Provability*, Cambridge, Cambridge University Press.

Bostock, D. (1997): *Intermediate Logic*, New York, Oxford University Press.

Bricker, P. (1987): 'Reducing Possible Worlds to Language', in *Philosophical Studies*, 52, 1987, pp. 331-355.

Carnap, R. (1947): *Meaning and Necessity, a Study in Semantics and Modal Logic*, Chicago, University of Chicago Press.

Chalmers, D. J. (2002a): 'On Sense and Intension', in *Philosophical Perspectives*, 16, pp. 135-82.

Chalmers, D. J. (2002b): 'The Components of Content', in Chalmers (2002c): pp. 608-633.

Chalmers, D. J. (ed.) (2002c): *Philosophy of Mind: Classical and Contemporary Readings*, New York, Oxford University Press.

Chalmers, D. J. (2003): 'The Nature of Narrow Content', in *Philosophical Issues*, 13, pp. 46-66.

Chalmers, D. J. (2004): 'Epistemic Two-Dimensional Semantics', in *Philosophical Studies*, 118, pp. 153-226.

Chalmers, D. J. (forthcoming): 'The Nature of Epistemic Space', in *Epistemic Modality*, Egan, A. & Weatherson, B. (eds.), Oxford University Press, forthcoming. Last accessed 16th Jan. 2010 on `http://consc.net/papers/espace.pdf`

Cherniak, C. (1986): *Minimal Rationality*, Cambridge, MIT Press.

Crimmins, M. & Perry J. (1989): 'The Prince and the Phone Booth: Reporting Puzzling Beliefs', in *Journal of Philosophy*, 86, pp. 685-711.

Cummins, R. & Pollock, J. (eds.) (1991): *Philosophy and AI: Essays at the Interface*, MIT Press, Cambridge.

D'Agostino, M. & Floridi, L. (2009): 'The Enduring Scandal of Deduction-Is Propositional Logic Really Uninformative?', in *Synthese*, 167, pp. 271-315.

DeRose, K. (1991): 'Epistemic Possibilities', in *Philosophical Review*, 100, pp. 581-605.

Divers, John (2002): *Possible Worlds*, London, Routledge.

Dretske, F. (1981): *Knowledge and the Flow of Information*, Cambridge, MIT

Press.

Duc, H. N. (1995): ‘Logical Omniscience vs. Logical Ignorance: On a Dilemma of Epistemic Logic’, in Mamede & Pereira (1995): pp. 237-248.

Duc, H. N. (1997): ‘Reasoning about Rational, but Not Logically Omniscient, Agents’, in *Journal of Logic and Computation*, 7, pp. 633-648.

Drapkin, J. & Perlis, D. (1986): ‘Step-logics: An Alternative Approach to Limited Reasoning’, in *Proceedings of the European Conference on Artificial Intelligence*, Brighton, England, pp. 160-163.

Eberle, R. A. (1974): ‘A Logic of Believing, Knowing, and Inferring’, in *Synthese*, 26, pp. 356-382.

Egan, A., Hawthorne, J., Weatherson, B. (2005): ‘Epistemic Modals in Context’, in Preyer & Peter (2005): pp. 131-170.

Egan, A. (2007): ‘Epistemic Modals, Relativism, and Assertion’, in *Philosophical Studies*, 133, pp. 1-22.

Elgot-Drapkin, J., Miller, M., Perlis, D. (1991): ‘Memory, Reason, and Time: The Step-Logic Approach’, in Cummins & Pollock (1991): pp. 79-103.

Fagin, R., Halpern, J. Y., Moses, Y., Vardi, M. Y. (1995): *Reasoning about Knowledge*, Cambridge, MIT Press.

Field, H. (2001): *Truth and the Absence of Fact*, Oxford, Oxford University Press.

Gabby, D. M. & Wansing, H. (eds.) (1999): *What is Negation*, Dordrect, Kluwer Academic Publishers.

Glüer, K. & Wikforss, Å. (2009): ‘Against Content Normativity’, in *Mind*, 118, pp. 31-70.

Hajek, P., Valdes-Villanueva, L., Westerstahl, D. (eds.) (2005): *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, Kings’ College Publications.

Hendricks, V., Jørgensen, K., Pedersen, S. A. (eds.) (2003): *Knowledge Con-*

tributors, Synthese Library, vol. 322, Kluwer Academic Publishers.

Hintikka, J. (1962): *Belief and Knowledge*, Ithaca, Cornell University Press.

Hintikka, J. (1969): *Models for Modalities*, Dordrecht, D. Reidel Publishing Company.

Hintikka, J. (1973): *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*, Oxford, Clarendon Press.

Hintikka, J. (1975): 'Impossible Possible Worlds Vindicated', in *Journal of Philosophical Logic*, 4, pp. 475-484.

Hintikka, J. (1989): *The Logic of Epistemology, and the Epistemology of Logic*, Dordrecht, Kluwer Academic Publishers.

Hintikka, J. (2003): 'A Second Generation Epistemic Logic and Its General Significance', in Hendricks et al. (2003): pp. 33-56.

Horwich, P. (1998): *Meaning*, New York, Oxford University Press.

Huemer, M. (2007): 'Epistemic Possibility', in *Synthese*, 156, pp. 119-142.

Humberstone, L. (2000): 'The Revival of Rejective Negation', in *Journal of Philosophical Logic*, 29, pp. 331-381.

Hunter, G. (1971): *Metalogic - An Introduction to the Metatheory of Standard First Order Logic*, London, Macmillan, 1971.

Jackson, F. (1998): *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford, Oxford University Press.

Jago, M. (2006): *Logics for Resource-Bounded Agents*, Ph.D. thesis, University of Nottingham.

Jago, M. (2009a): 'Logical Information and Epistemic Space', in *Synthese*, 167, pp. 327-341.

Jago, M. (2009b): 'Epistemic Logic for Rule-Based Agents', in *Journal of Logic, Language and Information*, 18, pp. 131-158.

Jeffrey, R. (1983): *The Logic of Decision*, Chicago, University of Chicago Press.

Konolige, K. (1986): *A Deduction Model of Belief*, Morgan Kaufman.

Lemmon, E. J. (1998): *Beginning Logic*, ninth printing, Indianapolis, Hackett Publishing Company.

Lewis, D. (1986): *On the Plurality of Worlds*, Oxford, Blackwell Publishers.

Lewis, D. (2004): ‘Letters to Priest and Beall’, in Armour-Garb et al. (2004): pp. 176-177.

Lycan, W. (1990): ‘Mental Content in Linguistic Form’, in *Philosophical Studies*, 58, pp. 147-154.

Mamede, N. & Pereira, C. P. (eds.) (1995): *Progress in Artificial Intelligence. Proceedings of EPIA ’95*, vol. 990 of *Lecture Notes in Artificial Intelligence*, Heidelberg, Springer Verlag.

Mendelson, E. (1997): *Introduction to Mathematical Logic*, 4th edition, London, Chapman & Hall.

Nolan, D. (1996): ‘Recombination Unbound’, in *Philosophical Studies*, 84, pp. 239-262.

Nolan, D. (1997): ‘Impossible Worlds: A Modest Approach’, in *Notre Dame Journal of Formal Logic*, 38, pp. 535-572.

Pascal, E. (1998): ‘Believing, Holding True, and Accepting’, in *Philosophical Explorations*, 1, pp. 140-151.

Preyer, G. & Peter, G. (eds.) (2005): *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, Oxford, Oxford University Press.

Priest, G. (1999): ‘What Not? A Defence of a Dialetheic Theory of Negation’, in Gabby & Wansing (1999): pp. 101-120.

Priest, G. (2005): *Towards Non-Being: the Logic and Metaphysics of Intentionality*, Oxford, Oxford University Press.

Priest, G. (2006a): *Doubt Truth to be a Liar*, Oxford, Oxford University Press.

Priest, G. (2006b): *In Contradiction: A Study of the Transconsistent*, second edition, Oxford, Oxford University Press.

Rantala, V. (1975): ‘Urn Models: A New Kind of Non-Standard Model for

First-Order Logic', in *Journal of Philosophical Logic*, 4, pp. 455-474.

Restall, G. (2005a): 'Multiple Conclusions', in Hajek et al. (2005): pp. 189-205.

Restall, G. (2005b): 'Logics, Situations and Channels', in *Journal of Cognitive Science*, 6, pp. 125-150.

Robbins, P. (2004): 'To Structure, or Not to Structure?', in *Synthese*, 139, pp. 55-80.

Rumfitt, I. (2000): '"Yes" and No"', in *Mind*, 109, pp. 781-823.

Schotch, P., Jensen, J., Larsen, P., MacLellan, E. (1978): 'A Note on Three-Valued Modal Logic', in *Notre Dame Journal of Formal Logic*, 19, pp. 63-68.

Sequoiah-Grayson, S. (2008): 'The Scandal of Deduction—Hintikka on the Information Yield of Deductive Inferences', in *Journal of Philosophical Logic*, 37, pp. 67-94.

Smiley, T. (1996): 'Rejection', in *Analysis*, 56, pp. 1-9.

Soames, S. (1987): 'Direct Reference, Propositional Attitudes and Semantic Content', in *Philosophical Topics*, 15, pp. 44-87.

Soames, S. (2005): 'Kripke on Epistemic and Metaphysical Possibility: Two Routes to the Necessary Aposteriori', in *Saul Kripke*, Berger, A. (ed.), Cambridge, Cambridge University Press, forthcoming. Last accessed 16th Jan. 2010 on `http://www-rcf.usc.edu/~soames/forthcoming_papers/Kripke_on_Epistemic.pdf`

Speaks, J. (2006): 'Is Mental Content Prior to Linguistic Meaning?', in *Nous*, 40, pp. 428-467.

Stalnaker, R. (1984): *Inquiry*, Cambridge, MIT Press.

Stanley, J. (2005): 'Fallibilism and Concessive Knowledge Attributions', in *Analysis*, 65, pp. 126-131.

Teller, P. (1972): 'Epistemic Possibility', in *Philosophia*, 2, pp. 303-320.

Vander Laan, D. (1997): ‘The Ontology of Impossible Worlds’, in *Notre Dame Journal of Formal Logic*, 38, pp. 597-620.

Wansing, H. (1990): ‘A General Possible Worlds Framework for Reasoning about Knowledge and Belief’, in *Studia Logica*, 49, pp. 523-39.

Wassermann, R. (1999): ‘Resource Bounded Belief Revision’, in *Erkenntnis*, 50, pp. 429-446.

Wedgwood, R. (1999): ‘The A Priori Rules of Rationality’, in *Philosophy and Phenomenological Research*, 59, pp. 113-131.

Yagisawa, T. (1988): ‘Beyond Possible Worlds’, in *Philosophical Studies*, 53, pp. 175-204.

Zalta, E. (1997): ‘A Classically-Based Theory of Impossible Worlds’, in *Notre Dame Journal of Formal Logic*, 38, pp. 640-660.