MODELS OF PROPOSITIONAL CONTENT

BY

ERIC ALAN MACK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Associate Professor Daniel Korman, Chair
Professor Timothy McCarthy
Professor Peter Lasersohn
Lecturer Brian Rabern, University of Ediburgh

ABSTRACT

Propositions, in addition to being the things that sentences express relative to contexts of utterance, can be invoked to play a few theoretical roles: since a sentence seems to be true just in case it expresss a true proposition, propositions could be seen the primary bearers of truth and falsity; since sentences can be said to be *necessarily* or *possibly* true in virtue of expressing necessarily or possibly true propositions, propositions could be seen as being the primary bearers of modal properties; since it is possible to know what someone said, propositions would be the things that we are properly said to *know*. Propositions are of deep philosophical interest mostly due to the fact that each of these four theoretical roles involves a perennial philosophical subject—meaning, truth, modality, and knowledge. There should be no surprise that philosophers are intent on analyzing and understanding propositions.

Of course, merely specifying a list of philosophically interesting theoretical roles does not suffice as an analysis. Other than the fact that propositions play these theoretical roles, they seem to be mysterious place-holders. What precisely are the things that can be simultaneously expressed by a sentence, true, necessary, and known? One ambition of a theory of propositional content is to point to a class of entities that can be modeled in a way that satisfies the philosophical demands of each of these roles. To this end, I provide a novel theory of propositional content and show that it yields solutions to problems plaguing its competitors. Based on a generalization of standard intensional models, I develop a formal framework in which propositions are identified with partitions of sets of possible worlds.

ACKNOWLEDGMENTS

I am very much indebted to a large set of individuals, without whom this project would have been a fantastic failure. In particular, I would like to thank my wife Abby for all the encouragement and support one could possibly hope for (and our dog Ike for quietly keeping me company whilst philosophizing). Thanks to my committee chair Daniel Korman, a model dissertation supervisor, for all the guidance and wisdom throughout my graduate studies. Thanks to my committee members Brain Rabern, Timothy McCarthy, and Peter Lasersohn for providing penetrating criticism while still encouraging creative ideas. Thanks to Allen Renear and the GSLIS conceptual foundations group, Jonathan Livengood, Timothy Cleveland, Zach Horne, and Steven Lee; all of whom provided great company and invaluable intellectual conversation. Finally, thanks to my family—parents, grandparents, brothers, and sister-in-law—for all the support.

Contents

Chapter 1

INDIVIDUATING PROPOSITIONS

It isn't that they can't see the solution.
It is that they can't see the problem.

Gilbert Keith Chesterton, 1874—1936

1.0

In this chapter we explore various philosophical motivations and prospects for a theory of propositional content.

1.1 Theoretical Roles

Propositions, in addition to being the things that sentences express relative to contexts of utterance, can be invoked to play a few theoretical roles: since a sentence seems to be true just in case it expresss a true proposition, propositions could be seen the primary bearers of truth and falsity; since sentences can be said to be *necessar-*

ily or *possibly* true in virtue of expressing necessarily or possibly true propositions, propositions could be seen as being the primary bearers of modal properties; since it is possible to know what someone said, propositions would be the things that we are properly said to *know*. Propositions are of deep philosophical interest mostly due to the fact that each of these four theoretical roles involves a perennial philosophical subject—meaning, truth, modality, and knowledge. There should be no surprise that philosophers are intent on analyzing and understanding propositions.

Of course, merely specifying a list of philosophically interesting theoretical roles does not suffice as an analysis. Other than the fact that propositions play these theoretical roles, they seem to be mysterious place-holders. What precisely are the things that can be simultaneously expressed by a sentence, true, necessary, and known? One ambition of a theory of propositional content is to point to a class of entities that can be modeled in a way that satisfies the philosophical demands of each of these roles. For illustration, suppose these entities were just sentences. Then, the existence of propositions would be no more controversial or mysterious than the existence of sentences. Unfortunately, this proposal fails to meet a basic constraint on a theory of meaning: sometimes, relative to contexts, distinct sentences can *mean* same thing. The sentences 'Bill likes Sue' and 'Sue is liked by Bill', for instance, express the same propositional content. If propositions were sentences, there would seem to be no way to account for this fact unless the identity relation for sentences depends on some prior commitment to propositional content—i.e. 'Bill likes Sue' and 'Sue is liked by Bill' are identical sentences (despite their surface appearance) in virtue of expressing the same proposition. But then, identifying propositions with sentences does not eliminate any mystery about the nature of propositions. Rather, sentences come to be more mysterious than we might previously have thought.[1]

The question 'what is the nature of propositions?' is very general and does not suggest any obvious method for finding a satisfactory answer. But the fact that there seems to be a many-one relation between sentences and propositional contents raises a somewhat more precise question: under what conditions do distinct sentences express the *same* proposition? This is a question about the identity conditions of propositions. In the same way that we would like to know the conditions under which a person at time t_1 is the same as a person at time t_2, we would like to know the conditions under which a proposition expressed by sentence s_1 is identical to the proposition expressed by sentence s_2. The question of trans-sentential identity of propositions is a useful starting point for a metaphysical understanding of propositions in the same way that the question of trans-temporal identity conditions for persons is a useful starting point for a metaphysical understanding of persons.[2]

Propositions induce some equivalence relation (that is not simply identity) on the set of sentences, and this fact suggests a useful dataset against which to evaluate our metaphysical theories of propositions—namely our intuitive judgments about the conditions under which distinct sentences express the same proposition. Any adequate theory of propositions should respect, if not *predict*, our intuitive judgments about sameness of meaning across sentences.

However, even though the question of cross-sentential sameness of meaning is a useful entering wedge for an adequate metaphysical theory of propositions, it fea-

[1]Quine (1986, chapter 1) provides an inductive argument for letting sentences play the role of propositions. Rather than belaboring the problems with this proposal, the first two chapters here will focus on reasons to think that the inductive argument is not cogent.

[2]Moreover, if Quine (1969, 23) is correct there is "no entity without identity", intelligible identity conditions are perhaps more important than merely providing a useful starting point for metaphysical understanding.

tures the quasi-technical notion of meaning at center stage. In challenging cases, our judgments tend to be theory-driven, as they are in challenging cases about personal identity over time. So, one methodological desideratum is to find ways of approaching the sameness of meaning question indirectly, by discussing various approximations of meaning. Instead of talking directly about sameness of meaning, we can talk directly about necessary conditions for sameness of meaning. Moreover, we can extract these necessary conditions from the other theoretical roles usually assigned to propositions. Consider, for example, the claim that propositions are the primary bearers of truth and falsity. We can translate this into a necessary condition for sameness of propositional content across sentences: two sentences express the same propositional content relative to a context only if they have the same truth value. This condition is trivial, but it is a start. For example, any theory that results in ‘Obama is President’ and ‘Biden is President’ expressing the same proposition is disqualified because it fails this trivial necessary condition for sameness of meaning.

Similarly, consider, the claim that propositions are the primary bearers of modal properties like necessary truth and possible truth. This theoretical role suggests the following necessary condition for sameness of propositional content across sentences: two sentences express the same propositional content (relative to a context) only if they have the same modal profile. Hence any theory according to which ‘chordates are things with a heart’ expresses the same propositional content as ‘renates are things with a heart’ is disqualified. The question of whether sameness of modal profile constitutes a necessary and *sufficient* condition for sameness of propositional content will take up a large portion of the present chapter, but for now the point is that it can be wielded as at least a necessary condition.

If a theory that is supposed to respect, if not predict, our judgments about the sameness of meaning across sentences is not disqualified by these two necessary conditions—sameness of truth value and sameness of modal profile—then it might still be ruled out by a further constraint. Consider the claim that propositions are the proper objects of knowledge. Then, of course, they are the objects of belief, assuming one must believe what one knows.[3] This theoretical role translates into the following powerful necessary condition for sameness of propositional content across sentences: two sentences express the same propositional content (relative to a context) only if it is impossible to believe one without also believing the other.[4] At face value, this condition is even stronger than the previous condition; it seems that two sentences have have the same modal profile while it is possible to believe one without believing the other—this question will also take up a significant portion of this chapter.

It might be objected that the original formulation of the sameness of propositional content question—'under what conditions do two sentences express the same thing?'—is really the same as the question of when two sentences are interchangeable in the complements of belief ascriptions, the latter is just the former in disguise. But this is not quite correct. First of all, there is an important methodological difference between the two. In considering relations between belief ascriptions, rather than relations between sentence contents, one is able to leverage two things: (i) The question becomes one about strict implications between sentences—i.e. necessarily, 'S believes

[3]The assumption that knowledge entails belief will not be too important for our purposes until chapter 4. Although it has a canonical status in epistemology, see Murray et al. (2013) for some reasons to think this entailment doesn't hold.

[4]Note that the translation of each of the three theoretical roles for propositions into their corresponding necessary conditions for sameness of propositional content depends implicitly on Leibniz' law. Hence, if P is true, and $P = Q$ then Q must also be true; if P is *necessarily* true, and $P = Q$, then Q must also be necessarily true; if an agent S believes that P and $P = Q$, then S believes Q.

that P' is true just in case 'S believes that Q' is true—which leverages intuitions about *what* necessarily entails *what*, rather than asking directly about whether two sentences have identical meanings.[5] (ii) Thinking about strict entailment patterns between belief ascriptions leverages an independent stock of intuitions about mental attitudes.

Secondly, interchangeability of sentences salva veritate in belief contexts is not obviously sufficient for sameness of propositional content. It is worth pointing out that, for Frege, strict entailment between belief ascriptions was a necessary *and* a sufficient condition for sameness of meaning.[6] The necessity claim, assuming that propositions are the proper objects of belief, is just an instance of Leibniz's law. The sufficiency claim is more substantive. Intuitively there *could* be pairs of sentences expressing distinct propositions which are such that no possible epistemic agent could believe one and not the other (e.g 'I exist' and 'Something exists'). For all intents and purposes, we will be operating as if Frege were correct about interchangeability in belief contexts being sufficient for sameness of propositional contents across distinct sentences. For the most part, Frege's criterion predicts the intuitively correct results, so it is useful for gauging whether or not a theory is on the right track.

The remainder of the present chapter is dedicated to the question of whether sameness of modal profile is a sufficient, rather than merely necessary, condition for sameness of propositional content. In particular, if the interchangeability of two sentences in belief contexts is a necessary condition for sameness of propositional content,

[5] Kaplan (1999, 33:30) points out that most of us tend to have more stable intuitions about logical consequence than about truth.

[6] "Now two sentences A and B can stand in such a [synonymy] relation that anyone who recognizes the content of A as true must thereby also recognize the content of B as true and, conversely, that anyone who accepts the content of B must straightaway accept that of A" Frege (1906).

doesn't this present a conflict for the claim that sameness of modal profile suffices for sameness of propositional content? The goal of the present chapter is to argue for an affirmative answer to this question, but also to argue that problems associated with the modal profile view—in particular, the problems of logical omniscience, and problems having to do with direct reference—have been exaggerated in a way that covers over some important distinctions. With a clarification of the problems in view, we will be in a position to develop an alternative account that meets a well-defined set of constraints and desiderata. Before that, we will have a look at some philosophical motivations for identifying the proposition a sentence expresses with that sentence's modal profile, i.e. its distribution of truth values across the set of all possible worlds.

1.2 Possible Worlds Propositions

In the previous section, it was noted that one of the theoretical roles assigned to propositions is to be the primary bearers of modal properties—that is, the modal profile of a proposition determines the modal profile of any sentence expressing that proposition. This entails that two sentences express the same proposition only if those sentences have the same modal profile. If this were also a sufficient condition—i.e. any two sentences with the same modal profiles express the same proposition—this would suggest an outright identification of propositions with truth conditions: a proposition just is the set of conditions relative to which a sentence can be evaluated as true or false. Even more directly: a proposition is identical to a set of circumstances, or possible worlds. This is a familiar characterization of propositions, and the problems it raises are well-known.

But, recognizing that propositions have to be individuated at least truth-conditionally—

which is to say that two sentences cannot express the same proposition if they do not share truth values in all possible worlds—is a somewhat weak motivation for an outright identification between propositions and truth conditions. There are at least two other ways of motivating this identification. First, there is a sense in which truth-conditions, modeled as sets of possible worlds, are able to play at least *some* of the theoretical roles we have allocated for propositions. Consider the role of being the primary bearers of truth. A proposition, on the truth-conditional view, is true just in case our world is a member (in the set-theoretic sense) of that proposition. To say that the sentence 'Grass is green' is true is to say that the proposition that it expresses is true, which is to say that our world is a member of the set of worlds in which grass is green. Concerning modal properties, there is a familiar story to tell about what constitutes, for example, a proposition's being necessary: it contains all possible worlds. So, another motivation for identifying propositions with truth-conditions, modeled as sets of possible worlds, comes from the sense it which sets of possible worlds play some of the roles allocated to propositions.

There is a further principled motivation for identifying propositions with sets of possible worlds. Even if it is possible to make a convincing case that sets of possible worlds can be the primary bearers of truth and the primary bearers of modal properties, letting them be the objects of propositional attitudes seems, at least prima facie, to be more of a theoretical artifact than a happy prediction. But this might be due to an insufficiently clear understanding of what it is for an agent to hold an attitude toward a proposition, rather than some defect with the identification of propositions with sets of possible worlds.

Although not uncontroversial, we will suppose first of all that the propositional attitudes are relational and that the relata are agents and propositions—call this the *relational* account of propositions. According to the relational account, a belief ascription 'S believes that P' is true just in case the agent S stands in the belief relation to the propositional referent of the complement clause; a knowledge ascription 'S knows that P' is true just in case S stands in the knowledge relation to the propositional referent of the complement clause. The relational account is by no means the only option available for modeling propositional attitude ascriptions—alternatives to the relational account usually posit some intermediate vehicle of representation between the agent and the referent of the complement, but we will set these alternatives aside since the relational view is still alive and well in the propositional attitudes literature.[7]

Even if the relational account of propositional attitudes makes a non-trivial substantive claim about the truth conditions of propositional attitude ascriptions, it is still not very informative. At least prior to a clear characterization of propositional contents, the relation that obtains when an agent holds an attitude toward a proposition is somewhat mysterious. If propositions are non-mental, non-physical entities, then what is it for an agent to stand in, say, the belief relation to a proposition? Identifying propositions with sets of possible worlds seems to make the question even more challenging. How can an agent stand in an attitudinal relation to an abstract set of possibilities? Frege, for one, was comfortable with positing a grasping relation between epistemic agents and abstract senses expressed by sentences. Holding any propositional attitude, for Frege, entails first that the grasping relation obtains—so

[7] Although for an example of a non-relational theory of propositional attitude reports, see Moltmann (2003) which contains a defense of a Russellian "multiple-relation" theory of belief, originating in Russell (1912), Russell (1913), and Russell (1918).

in this sense grasping is seen to be fundamental among the attitudes.[8]

But the naturalistically inclined have found the Fregean grasping relation mysterious. Moreover, if Fregean senses are modeled with (or supplanted by) sets of possible worlds, it seems impossible to make sense of grasping in naturalistically respectable terms.

However, a causal account of propositional attitudes, whose chief defender is Robert Stalnaker, is surprising congenial for advocates of the possible worlds propositions. The so-called "causal-pragmatic" account aspires to characterize belief states and desire states as special instances of representational states. Here is Stalnaker's summary of the *causal-pragmatic* account of belief and desire in Stalnaker (1984, 7):

> The strategy I have in mind is the one suggested by the pragmatic picture of mental acts and attitudes. Belief and desire, the strategy suggests, are correlative dispositional states of a potentially rational agent. To desire that P is to be disposed to act in ways that would tend to bring it about that P in a world in which one's beliefs, whatever they are, were true. To believe that P is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which P (together with one's other beliefs) were true.

The causal-pragmatic account presupposes that representation—a relation common to all propositional attitude states—is a special kind of causal relation, involving a disposition to be in certain states in response to environmental conditions. For illustration, an agent's environment can be in one of n possible states (E_1, ..., E_n).

[8]In the following passage, Frege quite poetically describes the non-physical grasping relation: "[S]ince the answer lies in the non-sensible, perhaps something non-sensible could also lead us out of the inner world and enable us to grasp thoughts where no sense-impressions were involved. Outside one's inner world one would have to distinguish the proper outer world of sensible, perceptible things from the realm of the nonsensibly perceptible. We should need something non-sensible for the recognition of both realms but for the sensible perception of things we should need sense-impressions as well and these belong entirely to the inner world." Frege (1956, 309)

What is it for an agent to represent its environment? Under ideal conditions the agent can be in one of n corresponding states (O_1,...,O_n) which are counterfactually correlated to the respective environmental states—"for each i, the organism will be in state O_i if and only if the environment is in state E_i [where the] if and only if is causal [...]."[9] So, the representation relation featuring in Stalnaker's causal-pragmatic account is a relation of counterfactual correspondence. Representational mental states, like belief and desire, are thought of as special instances of the more general causal representational relation. The *pragmatic* aspect of the causal-pragmatic account is its characterization of belief in terms of desire, and vice versa.

Here are some observations about the causal-pragmatic representation relation. First, it makes sense to call the relation *representational* insofar as counterfactual correlation (causal or not) between the states of a system and corresponding states of its environment underlies a kind of indication relation in the following sense: If a system is disposed to be in one of n states depending on whether its environment is in a corresponding state, then that system's state indicates a fact about its environment. For illustration, the states of a barometer represent atmospheric pressure insofar as its states indicate—under certain idealized conditions—the surrounding atmospheric pressure. We take this instance of the indication to be reliable just in case the states are counterfactually correlated with corresponding environmental states of atmospheric pressure. On the causal-pragmatic account there is nothing special about representational states—they do not involve any particularly mysterious irreducibly mental aspect. In a concise statement: representation just is counterfactual correlation because counterfactual correlation involves indication. It should be noted

[9]See Stalnaker (1986, 114—115).

that the representation relation need not be manifested in observable behavior. A system can be disposed to take on states that are counterfactually dependent on its environment, even if those states are merely internal states.

How then do we get from the causal-pragmatic account of the propositional attitudes to the claim that propositions are identifiable with truth conditions? Consider again what the latter claim predicts about the individuation of propositions over distinct sentences: two sentences are true (relative to a context) in the same set of possible worlds if and only if those sentences express the same proposition. Since propositions certainly are not *more* coarsely individuated than truth conditions, the "only if" direction of this biconditional is what should derive motivation from the causal-pragmatic account of the attitudes. If anything, we would be inclined to think that truth conditions are too coarse to be identified with propositional contents. Indeed, the causal-pragmatic account of the attitudes suggests that propositions are no more fine-grained than truth conditions.

The reason the causal-pragmatic account suggests a coarse-grained individuation of propositions across sentences is the fact that it is formulated in terms of the causal relation of representation, which is itself a special case of the relation of indication—a non-hyperintensional relation. That is, for a representational system s, if 's indicates that P' is true, and P is truth-conditionally equivalent to Q, then 's indicates that Q' is also true. It is precisely the fact that indication is understood in terms of counterfactual causal correspondence with possible environmental states that it is a non-hyperintensional relation. If, as the causal-pragmatic account entails, belief is merely a special case of the more general indication relation, then belief itself would also be a non-hyperintensional relation. Hence, one of chief reasons for doubting that propositions are sets of possible worlds—the thought that belief must be hyperintensional—is false according to the causal-pragmatic account. Since our pre-theoretic judgments about entailment patterns between propositional attitude ascriptions is supposed to be the primary source of data undermining coarse-grained propositions, this evidence should either be set aside or explained away if the causal-pragmatic account is true.

So, there are plenty of philosophical motivations for the identification between propositions and truth-conditions modeled as sets of possible worlds.[10] Before considering the primary objections to this view of propositions, it will be useful to point out some of the theoretical desiderata that are also respected by the possible worlds account of propositions. First, possible worlds propositions are language-independent entities. The content of a sentence should not be contaminated with the syntactic structure of that sentence. After all, there is no guarantee that distinct sentences expressing the same proposition will share the relevant syntactic structure. More importantly, by involving linguistic structure in propositional contents the line between form/content distinction is lost. There could be pairs of individuals who believe the same content, but whose languages differ—perhaps one member of the pair is an animal which lacks anything resembling human language.[11]

In the following section two of the most familiar objections to possible worlds propositions are considered and refined. Some of the common alternatives to possible worlds propositions are considered in the following chapter, where it is shown that certain desiderata that are met by possible worlds propositions are sacrificed for the sake of solving these two problems. The reason for belaboring the motivations for

[10]Stalnaker (1976) grants that a reduction of possible worlds to propositions is possible, although, he argues, a reduction in the opposite direction is preferable.

[11]cf Chapter 2.

possible worlds propositions is, in part, to promote an alternative to the possible worlds account, which is derived from the same set of motivations, but nevertheless circumvents the two objections in the following section.

1.3 Some Problems

Despite the fact that possible worlds propositions seem to be recommended by both the theoretical roles allocated to propositions *and* the causal-pragmatic account of propositional attitudes, they are undermined by our intuitions concerning entailment patterns between pairs of propositional attitude ascriptions. The main objective for this section is to refine these objections in a way that reveals the philosophical commitments that must be preserved by any adequate solution.

1.3.1 Logical Omniscience

Propositional attitudes like belief and knowledge do not seem to be closed under strict entailment.[12] But if propositions are sets of possible worlds, and propositional attitude ascriptions are true just in case their subject stands in the appropriate attitude relation to the proposition expressed by the complement clause, then it follows necessarily that propositional attitudes are closed under strict entailment (at least if Leibniz's law applies to propositional attitude predicates). This is perhaps the most familiar instance of the problem of *logical omniscience*.[13] This quick formulation of

[12]We will say that a sentence S strictly entails a sentence S' just in case all possible worlds in which S is true are worlds in which S' is true.

[13]The "problem of logical omniscience" is a misnomer because, strictly speaking, the "logical" consequences of a proposition constitute only a proper subset of the strict entailments (truth-conditional consequences) of a proposition. Nevertheless, we'll use the label in a way that covers over this distinction.

the problem of logical omniscience applies not just to possible worlds propositions, but to any characterization of propositions such that they are individuated truth-conditionally. But other phenomena that fall under the same category *do* require possible worlds propositions. Here is a list of related problems, which all fall under the general category of problems of logical omniscience (using *belief* as our representative propositional attitude):[14]

(1) If 'S believes that P' is true, and P strictly entails Q, then 'S believes that Q' is true.

(2) If 'S believes that P' is true, and P and Q strictly entail each other, then 'S believes that Q' is true.

(3) If S holds any beliefs at all, S must believe all necessary truths.

(4) If 'S believes $\bot$' is true for any necessary falsity $\bot$, then, for all P 'S believes that P' is true.

In fact, (2) is the only instance of the problem of logical omniscience on this list that is independent of the possible-worlds model of propositions. Notice that (2)—(4) are all consequences of (1), which itself is derivable from the identification of possible worlds with truth conditions.

There are two methods that are typically used to derive (1) from possible worlds propositions. The first, and likely the most familiar derivation, comes from Jakko Hintikka's analysis of belief predicates $B_S(x)$ (representing "S believes that x" in which x is a sentential variable) as quantifiers over domains of possible worlds:[15]

[14]See Jago (2012) and Halpern and Pucella (2007) for a more thorough survey of various problems falling under the category of "logical omniscience" problems.

[15]This analysis was developed in Hintikka (1962) and generalized to cover other propositional attitudes in Hintikka (1969)

(Belief) For all S and P: $B_S(P) \Leftrightarrow P$ is true in all possible worlds compatible with S's beliefs.

It is not hard to see that whenever P strictly entails Q, believing that P entails believing that Q, according to (Belief). So, taking propositional attitude predicates to be Hintikka-style quantifiers over possible worlds means accepting all of (1)—(4).

But, stepping back a bit, even if one rejects Hintikka's quantificational analysis of belief, it is possible to show that, indeed, *any* relational account of belief in which the second relatum is taken to be a set of possible worlds must be closed under strict entailment. Note that the possible worlds proposition expressed by a conjunction is the set-theoretic intersection of the possible worlds propositions expressed by the conjuncts—a conjunction is true at a possible world just in case both of its conjuncts are true. Moreover, the belief predicate seems to distribute over conjunction in the following sense: if 'S believes that P *and* Q', then 'S believes that P' is true as well as 'S believes that Q.' Granting both of these assumptions, and supposing that P strictly entails Q, if 'S believes that P' is true, then so is 'S believes that P *and* Q'. By distributivity of belief, it follows that 'S believes that Q' is true. So, the problems of logical omniscience (1)—(4) are independent of the familiar Hintikka-style quantificational analysis of belief predicates. Strictly speaking, the only thing we *need* to presuppose is the possible worlds theory of propositions, the relational analysis of belief, that conjunction is set-theoretic intersection, and distributivity of belief over conjunction.[16]

It is easy to focus only on the prima facie unhappy consequences of (1)—(4), but there is also a sense in which the set-theoretic structure of propositions generates some

[16]Note that even if distributivity is rejected, the intermediate conclusion of the foregoing argument is damning enough—the fact that 'S believes that P' is true should not intuitively necessitate that 'S believes P *and* Q' is true for all strict entailments Q of P.

desirable predictions. For example, it seems necessary that belief in a conjunction '$A \wedge B$' entails a belief in the permuted conjunction '$B \wedge A$'. The fact that the truth conditions of both conjunctions are the same means that the possible worlds propositions expressed are the same—so a belief in one entails a belief in the other. Likewise, with disjunction. These trivial predictions do not necessarily follow from just any theory of propositions. This is one desirable output of maintaining a strict distinction between linguistic form and linguistic content, which, if nothing else, the possible worlds theory of propositions does exceedingly well.

On the other hand, however, there are plenty of unhappy consequences of (1)—(4). If S believes that grass is green, it does not intuitively follow that S therefore believes that either grass is green or Obama is the president—perhaps S has never heard of Obama, or is incapable of grasping the concept of *president* in the first place. But, according to (1), if S believes the former, then S believes the latter. Consider a prediction of (2): if S believes that grass is green, then S believes that either grass is green or Obama is president and Obama isn't president (since the latter is truth-conditionally equivalent to the sentence 'Grass is green'). Consider (3). Surely it is possible for an epistemic agent to fail to believe some necessary truths. More precisely, it is possible that there is a sentence ϕ that expresses a necessary truth, such that for some agent S, 'S believes that ϕ' is false. For instance, one's belief that grass is green does not seem to necessitate that one also believes that one also believes all theorems of Peano Arithmetic. (4) falsely predicts that one cannot make a logical error without thereby believing absolutely every proposition.

One possible response to these consequences of (1)—(4) is to point out that an agent's being disposed to honestly reject a particular *sentence* is not a foolproof guide to whether or not the agent really holds belief in question. That is, even if the following disquotation principle is true:

> **(Disquotation)** If a normal English speaker, on reflection, sincerely assents to 'P', then he believes that P.

the corresponding biconditional is not:

> **(Biconditional Disaquotation)** A normal English speaker who is not reticent will be disposed to sincere reflective assent to 'P' if and only if he believes that P.[17]

In order to see why, let's suppose that substituting proper names in a sentence does not affect the content that the sentence expresses. Richard (1983) presents a case in which sincere reflective rejection of a sentence would not necessitate a corresponding failure to believe: Sally is in a phone booth speaking to Bill, but does not realize that the person in the phone booth across the street is Bill. Sally sees a steamroller headed toward the phone booth across the street and waves to the person (Bill) in the phone both, communicating that she thinks he is in danger. Bill (speaking to Sally) says that the woman across the street believes that I am in danger. Sally infers that the woman across the street believes that Bill is in danger. At the same time, Sally is disposed to sincerely deny that *she* believes that Bill is in danger. But, Sally *is* the woman across the street waving to Bill. Hence, Sally is disposed to sincerely and reflectively reject a sentence ('I believe that Bill is in danger') the content of which she, by stipulation, believes. This case seems to show that (Biconditional Disquotation) is false.

The question, then, would be whether the apparent counterexamples to (1)—(4) depend implicitly on (Biconditional Disquotation). It would appear not. In all the

[17] This formulation is borrowed from Kripke (1979).

proposed counterexamples to (1)—(4), the failure of the relevant entailment patterns seems to be due to the contents of the constituents of the sentences in question. The reason it seems possible for one to believe the content of 'Grass is green' while not believing the content of 'Grass is green or Obama is president and he isn't president' has to be due to the fact that one sentence is simply about the greenness of grass, while the other is about the greenness of grass *and* the presidential status of Obama. There is no implicit assumption that the subjects of any of these belief ascriptions are disposed to any particular assent/dissent behavior. But that means that appealing to the falsity of (Biconditional Disquotation) is unhelpful with respect to these cases.

If the counterexamples to (1)—(4) all depend on the fact the entailed beliefs are intuitively *about* something other than what the original beliefs are about, then perhaps this could do some work in distinguishing the favorable consequences of (1)—(4) from the unfavorable ones. Consider a case in which we hold fixed whatever it is that a sentence is intuitively about, and instead pile on some arbitrary amount of logical complexity:

(5) a. The Earth is round.

b. Either the Earth is round and the Earth is not round, or the Earth is round.

For a case like this it is not hard to imagine an agent who accepts the sentence (5a) who rejects the sentence (5b). But it is implausible to think that an agent could believe the *content* of (5a) without thereby believing the content of (5b). In other words, with respect to (5a) and (5b), it is not hard to imagine cases that are candidate counterexamples to (Biconditional Disquotation)—an agent might well accept one while rejecting the other, but it does not thereby make us think that an agent can believe the content of one without believing the content of the other.

This raises a question. What is it about cases like (5a) and (5b)—pairs of sentences whose atomic constituents are identical—that explains why they are at least more easily interchangeable in belief contexts than are arbitrary pairs of logically equivalent sentences? One plausible explanation would be that since there is no difference in the atomic constituents between (5a) and (5b), there simply can be no difference in what the sentences are intuitively *about*.[18] Both (5a) and (5b) are about the roundness of the Earth. And, since there is also no change in truth conditions between (5a) and (5b), we find them to be much more plausibly interchangeable than just any old pair of logically equivalent sentences.

Why should we say that the sentences (5a) and (5b) really are intuitively about the same thing? After all, they are syntactically very distinct from one another: one contains a conjunction and a negation and a disjunction, while the other has no logical expressions. Here is why this is a bad reason to attribute distinct contents to (5a) and (5b). If (5a) and (5b) have distinct contents in virtue of their distinct logical structure, then it seems that there would have to be an explanation for cases in which syntactic structure is distinct but where the identity of content is virtually undeniable. Consider, for instance, the follow pair:

(6) a. Bill likes Sue.

b. Sue is liked by Bill.

But, if there can be syntactically distinct sentences that are identical in content, then there must be some explanation of why pairs (5a) and (5b) succeed in expressing

[18] In general we should at a qualification about context here—"there can be no difference in what the sentences are intuitively about, *relative to a context*." (5a) and (5b) do not have any paradigm contextually sensitive expressions, but the qualification is necessary for pairs of sentences whose atomic constituents contain contextually sensitive expressions.

distinct contents (that is, given that the reader does not share the intuition that (5a) and (5b) are interchangeable in belief contexts). Suppose, then, that the distinction (between pairs like (5a)/(5b) versus pairs like (6a)/(6b)) has to do with the fact that, while (6a) and (6b) have distinct syntactic structure, (5a) and (5b) have distinct *logical* structure. This sort of reasoning presupposes that there is a hard line to be drawn between differences in mere syntactic structure versus differences in logical structure, but this is not what's wrong with the reasoning. The problem is that even if one thinks of the logical constants as being special cases of so-called "syncategorematic" expressions—expressions having no individual meanings—then a change in logical structure introducing distinct logical constants (such as in the case of (5a) and (5b)) would seem to be incapable of inducing a change in content.[19] If this is correct, then it follows that introducing arbitrary logical complexity (such as in the case of (5a) and (5b)) preserves content, at least insofar as the new complexity does not change truth conditions. This explains the intuition one might have that arbitrarily complex logical reformulations of a simple sentence can be substituted in belief contexts salva veritate—the possible inappropriateness of actually doing so could be accounted for in Gricean terms.[20] Moreover, the falsity of (Biconditional Disquotation) would allow for the possibility of an agent rejecting an arbitrarily complex sentence, while nevertheless believing its content.

If these considerations are correct, then under the condition that truth conditions as well as the atomic constituents occurring in sentences are held fixed, the resulting logical omniscience phenomena are not actually problematic. The truly problematic

[19]See Peacocke (1976), McCarthy (1981), and MacFarlane (2000) for various ways of understanding the claim that logical constants are *content neutral*.

[20]It is not difficult to construct an explanation of this pragmatic inappropriateness in terms of flouting the maxim *manner* in Grice (1975).

cases arise from logical manipulations that change the set of atomic constituents of a given sentence—even if such manipulations preserve truth conditions. The problems of logical omniscience are genuine problems in need of a solution, but the point here is merely to clarify the problem in order to open the door to a candidate solution.

1.3.2 The Frege-Soames Puzzle

So much for the first perennial challenge to possible worlds propositions. In this section another equally compelling, although somewhat less familiar, challenge to possible worlds propositions is explored. The supposition that the semantic value of proper names and indexicals are their referents leads to a well-known version of Frege's puzzle, which involves entailment patterns between propositional attitude ascriptions. Consider the following principle of compositionality:

> **(Comp)** If S_1 and S_2 are non-intensional sentences or formulas with the same grammatical structure, which differ only in the substitution of constituents with the same semantic contents (relative to their respective contexts and assignments), then the semantic contents of S_1 and S_2 will be the same (relative to those contexts and assignments).

For co-referential names or indexical expressions e and e^*, (7a) and (7b) should be equivalent for all instances of S and P:

(7) a. S believes that $P(e)$.

b. S believes that $P(e^*)$.

But there are instances of S and P for which, intuitively, (7a) is true and (7b) is false. For example, 'Louis believes that Superman in strong' is true, but 'Louis believes that Clark Kent is strong' is false. Note that this version of Frege's puzzle does not depend on any commitment to possible worlds propositions. And for those who find this version unproblematic, Soames (1987) provides a strengthened version of it aimed particularly at possible worlds propositions—two premises of which we have already been assuming: (i) that the semantic contents of names and indexicals are their referents, and (ii) that a relational account of propositional attitude ascriptions is true. In addition, assume that (iii) the semantic content of a declarative sentence is a set of truth-supporting circumstances (possible worlds) in which that sentence is true, and (iv) that belief distributes over conjunction (as understood in section 3.1 above). Finally, assume that the following is constitutive of conjunction:

> **(Conj)** The set of possible worlds in which a conjunction is true is the intersection of the sets in which each of its conjuncts are true.

Now, consider the following reductio of these assumptions:

(8) a. The ancients believed that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus. (assumed)

b. The ancients believed that 'Hesperus' refers to Phosphorus and 'Phosphorus' refers to Phosphorus. (by (i) and (Comp))

c. The ancients believed that 'Hesperus' refers to Phosphorus and 'Phosphorus' refers to Phosphorus, and there is an x such that 'Hesperus' refers to x and 'Phosphorus' refers to x. (by (ii) and (Conj))

d. The ancients believed that there is an x such that 'Hesperus' refers to x and 'Phosphorus' refers to x. (by (iv))

Exactly what this reductio counts as a reductio *of* is not so clear. If one finds the first version of Frege's puzzle (the equivalence of all instances of (7a) and (7b)) sufficiently compelling, then (8a)—(8d) might well be seen as confirmation that (i) must be false. But since (i) is well-supported by independent Kripkean considerations,[21] the culprit is plausibly something else. Moreover, even if (i) is to blame, it is possible to formulate a version of (8a)—(8d) using replacing (i) with the somewhat less contentious claim that the semantic values of indexical referring expressions (relative to contexts) are their referents.[22]

The argument (8a)—(8d) is supposed to illustrate that rigidly designating proper names, while they might well lead to Frege-style puzzles, present a *special* challenge for defenders of possible worlds propositions. After all, (8b) follows from (8a) without appealing to possible worlds propositions—so if this inference is problematic, it is problematic for anyone holding that 'Hesperus' and 'Phosphorus' rigidly designate the same object. However, notice that there is a plausible reading of (8b) whereby, if (8a) is true, so is (8b). Suppose that the speaker and hearer of (8a) are both aware that 'Hesperus' and 'Phosphorus' both denote Venus. It seems acceptable—or at least *plausible*—to infer (8b) from (8a). Notice, however, that there is no such plausible reading of (8d). According to the relevant reading (8b), it does not follow that the ancients believed there is one object to which 'Hesperus' and 'Phosphorus' refer. This seems to indicate that, even if rigid designation is problematic for everyone, it is *especially* problematic for defenders of possible worlds propositions.

One possible option for blocking this reductio argument is to deny that belief dis-

[21]Considerations derived primarily from Kripke (1972).

[22]However, see Elbourne (2010) for some reasons to think that indexical referring expressions have descriptive content.

tributes over conjunction, in which case it would be possible to believe a conjunction without believing both conjuncts individually. If this were the case, the inference from (8c) to (8d) would be invalid. So what's wrong with this option? First of all, it is simply implausible to suggest that an S could believe a conjunction $\phi \wedge \phi$ while failing to believe either ϕ or ψ. It seems to be constitutive of conjunction that belief distributes over it. If an agent accepts the sentence $\phi \wedge \psi$ but fails to belief either ϕ or ψ individually, then it is more plausible to infer that this is a counterexample to (Disquotation) than a case that shows that belief fails to distribute over conjunction. What about the familiar cases of the preface and the lottery that seem to show that belief fails to distribute over conjunction? Since this is not the space to evaluate these cases in any serious way, let's grant that it is at least plausible that belief fails to distribute over conjunction, so that the inference from (8c) to (8d) is plausibly the culprit in the reductio (8a)—(8d). The problem is that (8c), as Soames points out, is bad enough of a sub-conclusion on its own. So, even granting that belief fails to distribute over conjunction, the reductio (8a)—(8c) still stands.

Another possible option is to deny (Comp), which is supposed to be relatively uncontroversial in the context of the other assumptions behind the reductio (8a)—(8d). This requires making a somewhat ad hoc distinction between truth-supporting circumstances, on the one hand, and worlds on the other. While it might well be constitutive of '$\wedge$' that $\phi \wedge \psi$ is true in exactly those worlds in which ϕ is true and ψ is true, this need not be true of all indices relative to which sentences are evaluated to truth or falsity. At this point, this sort of solution seems ad hoc and unmotivated. In chapter 3, I will present a theoretically-motivated distinction of the sort that is needed for denying (Comp). If successful, this would show that there is no obvious reason to think that rigidly designating names present a special challenge to propositions qua sets of truth-supporting circumstances.

1.4 Summary

We have established two distinct challenges to possible worlds propositions. The philosophical motivation for identifying propositions with sets of possible worlds will ultimately be shown to suggest a way out of these challenges. Before that, in the following chapter we will look at two alternatives to possible worlds propositions and their respective obstacles. This will help to compile a collection of constraints on an adequate solution to the challenges raised in the present chapter.

Chapter 2

ADEQUACY CONSTRAINTS

2.0

In the last chapter, a formal theory of propositions was introduced along with its philosophical motivations. Two distinct problems for it—the logical omniscience phenomena and the Frege-Soames puzzle—seem to call for a more or less fundamental revision of the theory. In this chapter we briefly explore two leading alternatives to possible worlds propositions. This will serve both to situate the problems in a larger philosophical context and to further motivate a solution to the two problems raised in the last chapter. The correct solution to these problems should preserve the merits of possible worlds propositions over competing accounts of propositions.

2.1 Impossibilities

The first of the competing accounts of propositions to be explored is a generalization of the possible worlds account. According to the generalized version of the possible worlds account—endorsed in various forms by Hintikka (1979), Barwise and Perry (1983), Jago (2012), Jago (2013), Krakauer (2013), and others—propositions are simply sets of *worlds*, possible or impossible.[1] We will call them *impossible worlds propositions*. We have been implicitly assuming thus far that worlds are logically possible states of affairs relative to which sentences are either true or false, but not both. An impossible world w is a world that is either *incomplete*—at least one sentence is neither true nor false relative to w—or *inconsistent*—at least one sentence is both true and false at w. Let f be a function such that for any world w, $f(w)$ is the set of sentences that are true relative to w. We can provide a corresponding syntactic characterization of impossible worlds as follows: A world w is said to be *negation complete* for all sentences ϕ, either ϕ or $\neg\phi$ is a member of w. A world w is said to be *syntactically consistent* if it does not contain both a sentence and its negation. A world is *impossible* if it is either negation incomplete or syntactically inconsistent (or both). The impossible worlds proposition expressed by a sentence S is the set of all worlds w such that S is a member of $f(w)$.

Notice, first of all, that impossible worlds propositions need not be as coarsely individuated as possible worlds propositions. For example, the proposition expressed by 'The Earth is flat' need not be identical to the proposition expressed by 'The Earth is flat, and Obama is both president and not president.' Why? There could be impossible worlds relative to which the former is true, but the latter is false. Since truth-conditionally equivalent sentences express identical propositions according to possible worlds propositions, these two sentences must express the same proposition. So, impossible worlds propositions are more finely individuated than possible worlds

[1] Impossible worlds also feature in theories of counterpossibles—counterfactuals with impossible antecedents—in Nolan et al. (1997) and Brogaard and Salerno (2013), among others.

propositions. That being the case, defenders of impossible worlds propositions are in a position to circumvent the following version of the problem of logical omniscience: since sentences with identical truth conditions need not express the same proposition, a relational account of propositional attitude ascriptions is saddled with the result that all pairwise truth-conditionally equivalent sentences are interchangeable in complement clauses. Other logical omniscience phenomena are taken care of in a similar way. Propositional attitude predicates that distribute over conjunction are not necessarily closed under logical consequence because, in general, even if ψ is a logical consequence of ϕ, it need not be the case that the impossible worlds proposition expressed by ϕ is identical to that expressed by $\psi \wedge \phi$. And so on for the other logical omniscience phenomena listed in the previous chapter.

This last observation also underlies the solution that impossible worlds offer to the Frege-Soames puzzle. The inference from premise (10a) to (10b) is valid only if the set of worlds supporting the truth of the existential generalization of a sentence is a superset of the set of worlds supporting the truth of the sentence itself. In particular, the existential generalization of 'The ancients believed that 'Hesperous' refers to Phosphorus and 'Phosphorus' refers to Phosphorus" is the sentence 'There is an x such that 'Hesperus' refers to x and 'Phosphorus' refers to x.' The set of possible worlds supporting the truth of the latter is a superset of the set of worlds supporting the truth of the former. But impossible worlds propositions need not reflect logical relations as set-theoretic relations. So, the impossible worlds proposition expressed by 'There is an x such that 'Hesperus' refers to x and 'Phosphorus' refers to x.' need not be a superset of 'The ancients believed that 'Hesperous' refers to Phosphorus and 'Phosphorus' refers to Phosphorus." That being the case, the intersection of both propositions need not be identical to the impossible worlds proposition expressed by the later. Since this is precisely what is supposed to enable the inference from (10b) to (10c), impossible worlds propositions offer some way out of the Frege-Soames puzzle.

Impossible worlds propositions *do* offer a way to deal with the puzzles regarding belief ascriptions, but the logical structure of impossible worlds has so far been underspecified. The reason they are useful for circumventing the present problems for the semantics of propositional attitude ascriptions owes to the fact that their logical structure is so underspecified. In order to make this clearer, consider the following specification of the logical structure of worlds (still using f as a function mapping worlds to sets of sentences).

(1) For any set of sentences Γ, there is a world w such that $f(w) = \Gamma$.

In less formal terms, for any set of sentences we could come up with, there is a world that supports the truth of all and only those sentences.[2] This way of specifying the truth-supporting behavior of worlds makes them maximally fine-grained. Moreover, since the truth-supporting behavior of worlds is so fine-grained, the impossible worlds propositions expressed by sentences are correspondingly fine-grained. Call them *maximally fine-grained* impossible worlds propositions, and call their members *maximally fine-grained worlds*. These propositions fail to reflect logical relations between sentences in the sense that implication relations do not manifest as the subset relation—otherwise this maximally fine-grained account of propositions would not be able to address the problems of logical omniscience. With this precise characterization of the truth-supporting behavior of worlds, we can at least evaluate the resulting extreme version of impossible worlds propositions.

[2]For a defense of this view of worlds, see Priest (2005).

Let ϕ and ψ be sentences. The impossible worlds propositions expressed by ϕ and ψ cannot, on the maximally fine-grained account, stand in a subset relation with one another. In general, there are no two sentences that, on the maximally fine-grained account, express the same impossible worlds proposition. If that is the case, at least one desideratum of a theory of propositions is not met—there are no distinct sentences expressing the same content, contra our intuitive judgments about sameness of meaning. Given any impossible worlds proposition, we can locate the unique sentence expressing that proposition because it will contain one world, the f value of which is a singleton. In this sense, the maximally fine-grained impossible worlds theory is much like a *sententialist* theory, according to which sentences simple express themselves. Why? According to both accounts, every sentence expresses its own unique content. The set-theoretic structure of impossible worlds propositions does no obvious work over and above the sentences themselves. Of course, if one's aim is to represent sententialism as an extension of the familiar possible worlds framework, then the collapse into sententialism is not itself a compelling critique.

However, sententialism faces some independently devastating challenges. First, even though we have been presupposing that there is a single object language under discussion, it is necessary to have an account of content according to which translation between languages is a matter of preserving content between languages. On the maximally fine-grained account, there is no sense to be made of the fact that the same content is expressible across distinct languages. And if so, 'Snow is white' expresses a content distinct from that of 'Schnie ist weiss.' A sententialist cannot appeal to a sameness-of-content relation in order to match up sentences across languages. It is constitutive of sententialism that distinct sentences express distinct contents, because they express themselves.

Second, sententialism has trouble accounting for the fact that non-linguistic animals seem capable of holding propositional attitudes. It is implausible to suggest that ascribing, say, a desire to an animal is to assert that a relation holds between an animal and a sentence.[3] Perhaps, then, the animal stands in a relation to a sentence of its own mental language. If so, it seems there are two possibilities: either agents capable of holding propositional attitudes are so in virtue of standing in attitudinal relations to a sentences of single mental language, or there are distinct mental languages across agents. But both of these options represent a departure from the sort of sententialism resembling maximally fine-grained impossible worlds propositions.

Alternatively, it might seem that the defender of maximally fine-grained impossible worlds propositions is in a perfect position to account for translation as content preservation. The proposal would be to pair up sentences according to sameness of content, and to let f map worlds to sentences is such a way that no sentence occurs in the f image of a world without its counterparts across distinct languages. For instance, for any world w, $f(w)$ would contain 'Snow is white' just in case it also contains 'Schnee ist weiss.' Accordingly, it is possible for distinct sentences to express the same semantic content. But if this constitutes an acceptable solution to the problem of translation across languages, why not suggest a similar move in order to coarsen up impossible worlds such that for any world w, $f(w)$ contains a sentence ϕ if an only if $f(w)$ contains any sentence that expresses the same content of ϕ?

[3]See Davidson (1982) for an argument that animals cannot genuinely hold propositional attitudes at all. Quine, on the other hand, provides the following reason for thinking that animals can stand in attitudinal relations to sentences: "We may treat a mouse's fear of a cat as his fearing true a certain English sentence. This is unnatural without being therefore wrong. It is a little like describing a prehistoric ocean current as clockwise" (Quine (1956, 186)).

Maximally fine-grained propositions would cease to be maximally fine-grained—and this is for the better, since maximally fine-grained propositions have counterintuitive consequences regarding the sameness of meaning relation and, therefore, relational accounts of propositional attitude ascriptions.

This way of coarsening up maximally fine-grained propositions is problematic, but perhaps not for an obvious reason. It may seem that any appeal to sameness of content in individuating the contents of sentences is viciously circular. In this case, the defender of impossible worlds propositions would be appealing to sameness of content in order to impose certain constraints on the function f mapping worlds to sets of sentences. But even if the circularity is not vicious, the question of the conditions under which distinct sentences express the same content is still left open. By rejecting (1), the defender of impossible worlds propositions is committed to specifying the logical structure of worlds such that it is not completely trivial as it is in (1). At the opposite end of the spectrum from maximally fine-grained propositions are the coarse-grained possible worlds propositions discussed in the previous chapter. It seems that it should be possible to find some appropriate middle ground between the extremely coarse and the extremely fine in a way that does not rest on an appeal to sameness of content.

One possibility for finding this middle ground is to let some consequence relation govern the logical structure of impossible worlds. Such a relation would, of course, have to be weaker than the classical consequence relation. The problem with this proposal is that any such logical consequence relation would also govern the govern the beliefs of ordinary epistemic agents in the following sense: If $\models_w$ were the consequence relation governing worlds, then, given that belief is a relation to between agents and propositions that distributes over conjunction, belief must be closed under the $\models_w$ relation. The reasons for thinking that this is an unacceptable result were explored in section 3.1 of the previous chapter.

Another possibility would be to let the logical structure of worlds depend on a more basic notion of epistemic possibility—a world is *epistemically possible for an agent* A just in case it is compatible with the agent's beliefs.[4] For example, before one learns that, say, the gravitational constant on Earth is 9.8 meters per second per second, a world in which the gravitational constant on Earth is 7.5 meters per second per second would be epistemically possible. For the purposes of talking about the logical structure of worlds, the relevant notion of epistemic possibility is not relative to any particular agent. We will say that w is *epistemically possible* if it is possible that for some agent A, w is epistemically possible for A. This approach is advantageous for defenders of impossible worlds propositions. First, it is compatible with fact that distinct sentences can express the same proposition. For example, since there are presumably no epistemically possible worlds in which 'Bill likes Sue' is true but 'Sue is liked by Bill' is false, these two sentences express the same impossible worlds proposition. Second, the logical structure of worlds is not specified directly in terms of sameness of content; rather, it is specified in terms of an independent notion of epistemic possibility.

The central objection to this approach is that it renders relational accounts of propositional attitude ascriptions more or less predictively impotent. If one were to inquire into the closure properties of belief predicates, for example, the answer would

[4] There is a straightforward sense in which the totality of an Agent's epistemically possible worlds represents the knowledge state of an agent: since each world is compatible with everything the agent knows, the totality of all such worlds encodes the totality of the agent's knowledge—see Hintikka (1962).

be given in terms of epistemically possible worlds—the definition of which simply *presupposes* that the closure properties for belief predicates are given. Since there is no procedure for determining substantive answers to questions about closure properties of belief, this approach to impossible worlds propositions is theoretically unmotivated. Coarse-grained possible worlds possible worlds propositions at least come with a procedure for determining which belief ascriptions for from others, even if its predictions are wrong.

To summarize, impossible worlds models of propositional content either predict extremely fine-grained propositions, or else they simply leave open key questions about the granularity of propositions. Maximally fine-grained impossible worlds—worlds whose logical structure is trivial—underly a theory of content that effectively collapses into a kind of sententialism about propositional content. While impossible worlds models of propositions built on a notion of epistemic possibility presuppose, rather than *predict*, closure properties of propositional attitude predicates.

The following question arises: is it possible to construct the set of epistemic possibilities in a way that does not make epistemic possibility an unanalyzed theoretically primitive notion? Why would it not be possible to strike a middle ground between coarsely individuated metaphysical possibilities vs maximally fine-grained worlds? It seems prima facie possible to model epistemic possibility to varying degrees of idealization. On the most radically ideal end of the spectrum are logically possible worlds, and on the other end of the spectrum are the maximally fine-grained worlds. The task, then, is to specify a plausible construction of epistemically possible worlds without any ineliminable appeal to epistemic notions.

How would such a construction proceed? At the very least, epistemically possible worlds should be such that the "obvious" logical truths are always supported. For example, sentences such as $\neg(P \wedge \neg P)$ and $P \leftrightarrow \neg\neg P$ should always be true in epistemically possible worlds. Why? These are the sorts of sentences expressing propositions that it is expected of moderately rational agents to believe simply in virtue of being rational epistemic agents (even if they refuse to accept the sentences themselves). Also, epistemically possible worlds should be such that the "blatant" logical falsehoods are never supported. For example, sentences such as $P \wedge \neg P$ and $P \leftrightarrow \neg P$ should never be true in epistemically possible worlds. These are the sorts of sentences expressing propositions that it is expected of moderately rational agents to rule out in virtue of being rational epistemic agents. Unfortunately for the defender of impossible worlds propositions, there is no middle ground between the coarse-grained and the fine-grained that can accommodate these sorts of constraints—as shown in Bjerring (2013) and Bjerring and Schwarz (2014). Here is the argument in condensed form:

> Assume worlds—whether possible or impossible—are [negation complete] in the sense that for every sentence they verify either it or its negation. Consider a world w that verifies some complex contradiction C of, say, classical propositional logic. Since C is a contradiction, there is a proof of $\neg C$. That is, there is a sequence of sentences $S_1, ..., S_n$, ending in $S_n = \neg C$, each member of which is either a simple tautology or derivable from one or two earlier elements in the sequence by a simple logical rule like modus ponens. Given that worlds are [negation complete], w contains either S_i or $\neg S_i$ for each element in the sequence $S_1, ..., S_n$. So there are exactly three possibilities for w: either (i) w verifies the negation of the simple tautology, or (ii) w verifies the premises of a simple logical rule as well as the negated conclusion, or (iii) w verifies both C and $\neg C$. In each case, w is a trivially inconsistent world by the standards of classical propositional logic. (Bjerring and Schwarz (2014, 6))

In somewhat simpler terms, if a world supports the truth of an impossibility C, then it supports the truth of very obvious and trivial impossibilities of the sort that should not be considered epistemically possible for minimally rational agents.[5] This is a disappointing result for the prospects of grounding a theory of content on the notion of epistemic possibility. It turns out that there is no way to define impossible worlds such that only classically contradictory sentences supported are the "non-obvious" contradictions.

The authors point out that the argument makes ineliminable use of the assumption that impossible worlds are negation complete—for any world w and sentence P, w supports P or w supports $\neg P$. So, it is tempting to think that in order to circumvent this result we should simply posit that among the impossible worlds there are negation incomplete worlds. But this seems to make the notion of epistemic possibility contingent on whether or not an epistemic agent as such must recognize bivalence. Again, here is the reasoning in Bjerring and Schwarz (2014, 6):

> [F]ailure of logical omniscience can hardly be reduced to skepticism about bivalence. Consider an agent who is certain that either S_i or $\neg S_i$ is true, for all members of the sequence $S_1, ..., S_n$. Worlds that verify neither S_i nor $\neg S_i$ should then not count as live possibilities for her: they are not maximally specific ways things might be. But even in that context, the complex contradiction C may be deemed possible by the agent, and its negation $\neg C$ may provide her with non-trivial information. This time, the worlds she rules out cannot be partial worlds, since those were already ruled out from the start. We are left with the original problem.

[5]This can be strengthened even further. Instead of invoking the idea of a "minimally rational" agent, we could just appeal to what is and is not possible to believe *full stop*. In that sense, an epistemically impossible world would be a world that supports the truth of sentences the contents of which are impossible to believe, by rational agents or *any other* epistemic agent.

This is a serious limitative result about the utility of impossible worlds for modeling information, knowledge, and belief. Worlds-based models of information, according to which an agents' gaining information amounts to eliminating epistemic possibilities,[6] are useful for modeling intentional notions like belief insofar as they can be specified in extensional, set-theoretic terms. So, the impossibility of carving out epistemic possibilities from the space of maximally fine-grained impossible worlds severely limits the use of impossible worlds for modeling information.

2.2 Structuralism

Proponents of structured propositions hold that propositions are non-linguistic entities with constituents that are structured in some way—usually in a way that reflects the syntactic trees of the sentences expressing those propositions. One motivation for the analysis of propositions as structured entities is the flexibility provided by syntactic structure for individuating propositions. Even if two different sentences are true under the same conditions, the propositions expressed by those sentences could be different as a result of their reflecting the different syntactic structures of those sentences.

According to a version of structured propositionalism whose current defenders include Soames, Salmon, and King—rooted primarily in Russell (1903)—the sentence 'John loves Sue' expresses the proposition represented by a triple containing the "Loves" relation and the individuals $\langle$John$\rangle$ and $\langle$Sue$\rangle$: $\langle$Loves, $\langle$John$\rangle$, $\langle$Sue$\rangle\rangle$.[7]

[6]See Dretske (2008) for a defense of the view that gaining information amounts to the elimination of possibilities.

[7]See, for example, Salmon (1986), Soames (1987), and King (2007) for contemporary defenses of russellian structured propositions.

Logical combinations of sentences are composed recursively. For illustration, 'John does not love Sue' expresses the proposition ⟨NEG ⟨Loves, ⟨John⟩, ⟨Sue⟩⟩⟩, and 'John loves Sue or Bill sleeps' expresses the proposition ⟨DISJ ⟨Loves ⟨John⟩,⟨Sue⟩⟩, ⟨⟨Bill⟩ Sleep⟩⟩, and so on.[8] A variant of the neo-Russellian approach is a view, developed and defended independently by Cresswell and Lewis, according to which propositions are structured intensions.[9] The proposition (structured intention) expressed by the sentence 'John loves Sue' is a tuple of intensions (functions from possible worlds to extensions). For example, the intension I_{Loves} of the expression 'Loves' is a function from possible worlds to pairs of individuals, the first member of which loves the second.

The technical difference between neo-Russellian approaches and Cresswell/Lewis approaches is important because individuation of propositions on both approaches depends on individuation of the constituents of propositions. While the constituents of the former are objects and properties, the constituents of the latter are functions from possible worlds to extensions. Under certain assumptions, it is possible for the two approaches to make different predictions concerning the granularity of propositions. For example, according to neo-Russellians, the sentence 'A is is trilateral' expresses the structured proposition represented by ⟨Trilateral,⟨A⟩⟩ where 'Trilateral' stands for the property of being trilateral, and 'A' stands for the object A. What about the structured proposition expressed by the sentence 'A is triangular'? Is it the same proposition as that expressed by 'A is trilateral'? One reason to think that they express distinct propositions is that it seems possible to be unaware that, necessarily, if something is triangular, then it is trilateral (and vice versa)—so the two sentences

[8]Here, 'NEG' and 'DISJ' denote the usual truth functions for negation and conjunction, respectively.
[9]Lewis (1972); Cresswell (1985)

do not seem to be interchangeable in the complements of belief ascriptions. Notice that the neo-Russellian is in a position to hold that the two sentences express distinct propositions. As long as 'Triangular' and 'Trilateral' stand for distinct properties, replacing one for the other as the constituent of a structured proposition results in a distinct structured proposition. On the other hand, defenders of structured intensions can recognize no distinction between what is expressed by 'A is triangular' vs 'A is trilateral.'

The concern is that the choice between the two options is not predicted by more fundamental commitments about the individuation of properties. The neo-Russellian is able to posit ad hoc distinctions between the semantic contents of primitive predicates on an as-needed basis. The fact that primitive properties are the constituents of neo-Russellian propositions is an both an advantage and a liability. It is advantageous insofar as the neo-Russellian has a winning strategy for drawing distinctions between intuitively distinct propositions. But this comes at the cost of having a theory that rests on unanalyzed properties, which seem to be at least as mysterious as propositions were in the first place. Advocates of structured intensions, on the other hand, seem to be stuck with the coarsely-individuated propositions, at least at the atomic level.

Advocates of structured propositions also sacrifice a few key advantages of possible worlds propositions. For example, since possible worlds propositions do not encode any facts about the syntactic structure of sentences expressing them, they uphold a clear distinction between form and content. In particular, defenders of possible worlds propositions are able to maintain a clear distinction between *linguistic* form and content. Of course, for a structured propositionalist, this would not be a compelling objection. After all, linguistic form, for structured propositionalists, simply *is* an aspect of linguistic content. But this raises a question about the extent to which linguistic form should determine distinctions between linguistic content. In the previous section, the fact that translation between natural languages is supposed to preserve content was used to undermine impossible worlds models of propositions. But this objection reemerges for structured propositions, although in a different way. If linguistic form is an aspect of linguistic content, then to what extent do different syntactic patterns across languages prevent translations that faithfully preserve content?

The worry is not that advocates of structured propositions must posit a distinction for every conceivable syntactic distinction; they are free to discard some syntactic distinctions as being semantically irrelevant. Rather, the concern is that there seems to be no non-circular way of preferring some syntactic distinctions over others such that some are semantically relevant while others are not. What is it in virtue of which some syntactic distinctions play a role in distinguishing contents, while other syntactic distinctions play no such role? An advocate of structured propositions seems to be in no position to answer this question. Moreover, recognizing some syntactic distinctions to be semantically relevant while holding others to be semantically irrelevant cannot simply be done on an as-needed basis. The same sort of question would arise—what is it in virtue of which certain syntactic distinctions and not others are semantically significant?

Another version of this objection applies independently of concerns about translation between languages. Consider, for example, the following pair of sentences:

(2) a. Grass is green and the Earth is round.

b. The Earth is round and grass is green.

Intuitively, (2a) and (2b) express the same propositional content, at least in ordinary contexts. So a structured propositionalist would have to say that the syntactic distinction between (2a) and (2b) is semantically insignificant—linguistic form does not contaminate linguistic content to the point of (2a) and (2b) being semantically distinct in ordinary contexts. But, if the justification for holding that (2a) and (2b) express the same content is that commuting conjunctions preserves content, this justification is not available to defenders of structured propositions.

Another advantage that unstructured possible worlds propositions have over structured propositions lies in the fact that there is a straightforward story to tell about how possible worlds propositions have truth-conditions. The members of possible worlds propositions just are the conditions relative to which (contextually-disambiguated) sentences can be evaluated to truth and falsity. That being the case, sets of possible worlds supporting the truth of a sentence ϕ encode precisely the truth conditions of ϕ. There is therefore no apparent mystery about how possible worlds propositions could be *representational* entities—they are representational insofar as they impose conditions on how the world must be. Contrast this with the structuralist's position. If a proposition is a structured complex consisting of objects, properties, and relations, what is it in virtue of which such propositions represent things? In particular, what is it in virtue of which such propositions have the truth conditions they have?

Another version of this question arises from the structuralist's attempt to account for the *unity* of structured propositions? Why, for example, does the sentence 'Bill walks' express the proposition represented as (3)

(3) $\langle$BILL,$\langle$WALK$\rangle\rangle$

rather than expressing the following structured complex?

(4) ⟨⟨WALKS⟩,BILL⟩

The fact that there seems to be no obvious way to decide between the two representations is problematic because, at least mathematically, (3) and (4) are distinct objects. If propositions are literally identified with ordered tuples, the choice between (3) and (4) makes a difference.

The point here is not to establish conclusively that these questions are impossible to answer. Rather, it is to show that in order to answer these questions some further theoretical machinery seems necessary. By contrast, possible worlds propositions do not raise the same sorts of questions. Notice, however, that a version of this challenge *does* apply to maximally fine-grained *impossible* worlds propositions. Recall from the previous section that maximally fine-grained impossible worlds propositions end up being as fine-grained as sentences themselves. How, then, do they come to be representational? One suggestion might be that they encode truth conditions in the same way as possible worlds propositions, but they also enable fine-grained distinctions between truth-conditionally equivalent contents. The trouble, then, is that these further distinctions seem to have nothing to do with their capacity to represent. Consider again, for example, (2a) and (2b). The advocate of maximally fine-grained impossible worlds propositions would either have to claim that these sentences express distinct representations, or that they do not. At least in ordinary contexts, the former option is completely implausible. But, on the other hand, if (2a) and (2b) do *not* express propositions that are representationally distinct, what justification could one possibly have for positing a distinction between the *contents* expressed by (2a) and (2b)?

2.3 Inadequacies

Both of these options for circumventing the problems of logical omniscience and the Frege-Soames puzzle come with their own respective challenges. Of course, impossible worlds propositions and structured propositions are also by no means the only alternatives to possible worlds propositions. Nevertheless, they are two of the most popular views in the contemporary literature and they serve to illustrate some of the distinct advantages offered by possible worlds propositions. We will close this short chapter by making some general methodological observations that further suggest steering clear of impossible worlds and structured propositions altogether.

To what extent should a model of propositional content generate predictions concerning entailments between belief ascriptions? There are at least three possible positions one might take here. The first is to say that propositional contents are simply not in the business of generating entailment patterns between belief ascriptions. Whether, say, belief is closed under believed entailment is not a matter that should be determined by the contents of belief ascriptions. To think otherwise is to conflate questions about the norms of rationality with questions about content. If this is correct, then unless formal semantics is a part of normative epistemology, formal models of propositional content seem to be unnecessary for strictly semantic purposes—we should instead opt for a kind of primitivism about propositions.[10]

The second answer one might give is that a semantic theory of belief ascriptions

[10] Jeff King criticizes this sort of primitivism about propositions because of a worry about how primitive propositions could come to have truth conditions: "I think that propositions do have constituents. This is mainly because I find the idea of "simple fine grained propositions", fine grained propositions without constituents or parts, mysterious. What would make such a simple proposition be about, say, Paris as opposed to Santa Monica? In virtue of what would it have the truth conditions it in fact enjoys? I cannot see that these questions have answers if propositions are held to be simple and fine grained" (King (2007, 6))

is uninteresting unless it *does* generate predictions about the doxastic behavior of rational agents. In fact, this is precisely what is unattractive about both maximally fine-grained impossible worlds propositions as well as primitivism about propositions. How do structured propositions look from this standpoint? It depends on the style of structuralism in question. Clearly neo-Russellian structuralism fails as a framework for generating predictions about rational belief. Structured propositions whose constituents are unanalyzed properties, relations, and objects do not stand in the right sorts of relations to one another so as to generate predictions about, say, rational inference and belief revision. Any entailment patterns between belief ascriptions that arise out of a neo-Russellian model of structured propositions will depend on the individuation of properties and relations, which a neo-Russellian theory as such does not account for in more basic terms. By contrast, consider the predictions generated by Lewis-Cresswell style structured intensions. Whenever $P(\mathrm{x})$ and $Q(\mathrm{x})$ are co-intensional predicates, replacing one for the other in any extensional context would be content-preserving because the resulting structured intension would be identical. In this respect, not surprisingly, structured intensions are similar to regular intensions. The question is whether the interchangeability of sentences like 'A is triangular' and 'A is trilateral' in the complements of belief ascriptions should count as a true prediction about rational belief.

The third answer one might give is that, *yes*, models of propositional content should be in the business of generating predictions about entailment patterns between belief ascriptions, but such entailment patterns are merely artifacts of a relational theory of belief together with a correct individuation of contents. This is the view to be defended in the chapter 3. The aim is to construct propositional contents that share in the benefits and motivations of the possible worlds theory—i.e. unstructured, non-primitive objects that encode truth conditions of sentences, which are related to each other in a way that generates entailment patterns for propositional attitude ascriptions—and corrects for the problems—i.e. overly coarse granularity, which leads to the problems of logical omniscience and the Frege-Soames puzzles.

2.4 Informal Individuation Criterion

Up to this point, most of the discussion of propositional content has focused on the individuation of contents across sentences. In the previous chapter this was shown to underlie a class of related problems facing possible worlds propositions. It should be clear from the present chapter that individuation also presents problems for the familiar alternatives to possible worlds propositions. What has not been established is that the individuation of propositions is a problem in itself independent of *any* particular formal construction. This position was most clearly defended in Quine (1986, ch 1) in which competing formal constructions of propositions were shown to fail to partition natural language sentences properly. But, instead of following Quine's inductive hypothesis that *no* proper individuation is forthcoming and, therefore, showing how we might do without propositions (and sameness of meaning) altogether, the goal in what follows is to construct propositions in a way that avoids the objections so far raised.

If a proposition can be expressed by distinct sentences at all, then, independent of any particular formal construction of propositions, we should expect syntactic criteria to decide questions regarding sameness of meaning. One subconclusion of the previous chapter was that the problems of logical omniscience, the problems related to the closure of propositional attitude predicates under logical consequence, are only really problematic in a subclass of cases. In particular, using the language of classical first-order logic as a toy model, the proposal was that if distinct logically equivalent sentences embed the same atomic sentences, we should be happy to say that they express the same proposition—the thought being that syntactic recombinations of the same atomic constituents using sound logical rules is not enough to induce a change in content. This criterion, proposed in the previous chapter as a necessary and sufficient condition for determining sameness of content, is inconsistent with the predictions of the various formal constructions sketched so far. But, as it stands, it is only a success condition for a formal construction of propositional contents; what is lacking is a formal construction of propositions predicting precisely this criterion of individuation, which is the aim of the following chapter.

Chapter 3

A NEW MODEL OF PROPOSITIONS

3.0

The previous chapters laid out a family of related problems facing unstructured propositions. These problems seem clearly to outweigh the benefits associated with the philosophical motivations for unstructured propositional contents. The aim of the present chapter is to present a revised account of propositional content that avoids these problems and which does not resort to imposing any syntactic structure on propositions. We start by providing some intuitive motivation for the revised account.

3.1 Tracking Theories and Modular Machines

Recall from the first chapter that *tracking* solutions to the problem of intentionality—the problem of giving a naturalistic account of intentional mental states—yield an unacceptably coarse individuation of propositions. Here is a quick summary: According to tracking accounts, the states of a system that, under ideal conditions, track states of the environment are bearers of information about the environment. For example, under the right conditions, the finitely-many states of a digital thermometer can bear information about the environment's temperature because the thermometer tracks the environment's temperature. That is, a digital thermometer is capable of being in states $(O_1, O_2, ..., O_n)$ that reflect, or *indicate*, corresponding environmental temperature states $(E_1, E_2, ..., E_n)$ through causal co-variation.[1] This picture leads to a coarse individuation of content precisely because the relation of indication between O_i's and E_i's is a non-hyperintensional relation—if a state O_i indicates E_i, and, necessarily, E_i obtains if and only if E_j obtains, then O_i must also indicate E_j. So, if intentional states like belief are special cases of *indicating* states, then representational states do not make hyperintensional distinctions. A central reason, perhaps *the* central reason, for positing distinctions between necessarily equivalent contents vanishes.

There is nothing particularly controversial about this account of how the states of a physical system can be said to indicate possible states of the environment and, thereby, can be said to bear information about the environment. What *is* controversial is the claim such states are genuinely intentional and that the things indicated by such states are propositional contents.[2] But this further step will not be relevant

[1]For various tracking accounts of intentionality, see Dretske (1981), Stalnaker (1984), Stalnaker (1986), and Fodor (1990).

[2]Apart from the fact that this account of intentional mental states yields an unacceptably coarse individuation of contents, proponents of "phenomenal" accounts of representation (see, for exam-

here; what will be relevant is the innocent causal tracking relation itself. The tracking relation between possible states of a system and possible states of the environment provides the basis for constructing propositional contents, although not in the way envisaged by some proponents of the tracking account of intentionality. The question of how to construct propositions from the causal tracking relation turns on the question of how to individuate the possible states of a complex system with component modules that track various distinct aspects of the environment.

Consider a system that tracks exactly one aspect of the environment by being in one of two possible states S_1 and S_2. As far as the system is concerned there are two possible states of the environment—E_1 and E_2. This can be true even if the aspect of the environment being tracked is not *really* binary—the system just keeps track of the environment in a coarse-grained way. To say that the system tracks the relevant environmental states is to say that, for $i = 1$ or $i = 2$, the system is in S_i if and only if the environment is in E_i, where the "if and only if" is causal. Of course, there can be another system whose role it is to track exactly one *other* aspect of the environment—it is capable of being in one of two states S_1' and S_2' which reflect environmental states E_1' and E_2', respectively. In general, for any possible aspect of the environment, we can imagine a possible system whose role it is to keep track of *that* aspect. Suppose now that each of these systems are component modules of one unified representational system. The unified system has different components whose role it is to keep track of distinct aspects of the environment. If the system has n component modules, then as far as the system is concerned at its full tracking capacity, there are 2^n possible states that the environment can be in. In other words,

ple, Mendelovici (2013) and Kriegel (2013)) argue that tracking accounts are incomplete because intentionality includes an irreducibly phenomenal component.

when all component modules are active, the system is capable of indicating that the environment is in one of 2^n distinct states. This is not to say that there are only 2^n distinct possible states that the environment can really be in. There will be more than n distinct aspects that contribute to composing possible environmental states. Also, the individual aspects that are tracked by components of the system might actually have continuum-many possible values. The fact that the representational system's components do their representing in binary effectively discretizes the system's range of possible representations.

From the fact that in our imagined system there are only 2^n possible configurations of n binary components when all components are active (and the corollary that the system is limited to representing 2^n possible environmental states while operating at full capacity), it does not follow that the system is only capable of being in a total of 2^n distinct states. One of the benefits of modular design is that subsystems are capable of operating independently of one another. So, in case half of the component modules composing the system fail, the system as a whole can still be operative.[3]. But the fact that our imagined system has this modular structure means that the number of possible states of the whole system has to be the sum of the distinct possible states that each subsystem can can be in. Whenever some number of modules are inactive, the resulting subsystem gives rise to a fresh set of possible states.

This answers the question of how to individuate the possible states of a representational system. If we individuate states based on the presupposition that the system is only doing its representing if all component modules are active, then we end up with a lower number of possible states. But, recognizing the fact that representa-

[3]One consequence is that modular systems are "future proof" in the sense that new modules can be imported to interact with existing systems

tional systems can track distinct aspects of the environment with n modules that operate independently of one another, the number of possible states that the system can be in at any given time is far greater than 2^n. This is a consequence of the fact that fine-grained states of the system (states of the system when operating at full capacity) are numerically distinct from the coarse-grained states of which the former are special cases.

We can make things more concrete by considering a physical system that is composed of four modules, each of which are binary indicators for four environmental variables—temperature, pressure, moisture, and wind speed. It is a very coarse-grained system in the sense that its components yield only 1 or 0, depending on the environment: the component module for tracking temperature (a binary thermometer) reads 1 or 0 depending on whether the environment is hot or cold, the component module for tracking pressure (a binary barometer) reads 1 or 0 depending on whether atmospheric pressure is high or low, and so on. In this simple four-component representational system, calculating the number of its possible distinct states is straightforward counting exercise, but it illustrates how our approach to individuating representational states generates large numbers of possible states from simple systems: If the thermometer, barometer, hygrometer, and anemometer components of the system are all active, then there are $2^4 = 16$ possible states that the system can be in. If only a subset of the components are active, there are more possibilities to count. There are a total of $\binom{4}{3} =$ four possible choices of three components, where each of the four choices can take on $2^3 =$ eight distinct configurations. So the total number of states compiled from all the three-component versions of our system is 12. There are a total of $\binom{4}{2} =$ six possible choices of two components where each of the six choices can take on 2^2 = four distinct configurations. So the total number of states compiled from all the two-component versions of our system is 24. Finally, there are $\binom{4}{1}$ = four possible choices of one component where each of the four choices can take on two distinct configurations for a total of eight possible states compiled from all the single-component versions of our system. Summing everything up shows that our simple system can be in one of $16 + 12 + 24 + 8 = 60$ possible representational states. If we generalize this counting procedure, the number of distinct possible states for a modular system with n distinct (binary) components is $\sum_{k=1}^{n} \binom{n}{k} 2^k$.

But this is not an essay on design techniques for systems engineering. How does this relate to the goal of constructing propositional contents? One way to proceed would be to identify the possible states of the environment that are represented by a system with n binary components with possible worlds. Propositions might then be identified with functions from such worlds to truth-values. The finest-grained distinctions that the system can make among possible worlds is determined by its number of independent binary components. There is no need to consider the distinctions that subsystems can draw among possibilities because special cases of coarse-grained states are not numerically distinct from the states of which they are special cases. For example, our four-component binary system can be in a state in which only three of its four modules are active—indicating that it is hot, low-pressure, and windy. Alternatively, if the anemometer becomes inactive for one reason or another, the system (because of its modular design) would indicate just that the environment is hot and low-pressure. In this case, the system would be in a numerically distinct representational state without representing a numerically distinct environmental state—the distinct state of the system just represents less information about the same environmental condition. A system with n binary components can distinguish at most 2^n numerically distinct possible environmental states—that is, possible worlds. So, a function from possible worlds—in this sense—to truth-values would be a function over a fixed domain containing exactly 2^n possible worlds. Of course, this is just the usual coarse-grained characterization of propositions as sets of possible worlds, the semantic consequences of which we set out to avoid.

The present proposal for constructing propositions from the tracking account of representation is to reorient the focus to the states of a representational (i.e. tracking) system *itself* rather than the objective possibilities represented by the system. This approach has the advantage of being able to leverage the fact that for any representational system, there are always strictly more representational states than there are states to be represented. In the following section this proposal will be illustrated explicitly by applying it to a toy fragment of natural language, the quantifier/variable/identity-free fragment of a first-order language with logically independent atomic sentences. Even though this language is trivially simple, it exhibits the problematic semantic features of natural language that were highlighted in the previous two chapters. In particular, if it is interpreted with a standard possible worlds semantic model, then logically equivalent sentences express the same proposition, there is only one necessarily true proposition; there is only one necessarily false proposition; propositional operators (that distribute over conjunction) would be closed under logical entailment; the Frege-Soames puzzle for worlds-based propositions applies.[4]

[4]This list is not meant to be exhaustive; these were the issues discussed in the previous two chapters.

3.2 A Formal Model

Let $\mathcal{L}$ be a language whose alphabet consists of at most countably-many individual constants and n-ary predicate symbols for each natural number n, the usual logical connectives, and parentheses. The members of $\mathcal{L}$ are the atomic sentences $Pa_1...a_n$ (where P an n-ary predicate symbol) and complex sentences: $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, and $(\phi \rightarrow \psi)$, whenever ϕ and ψ are members of $\mathcal{L}$. So, our toy language $\mathcal{L}$ is just the propositional fragment of the language of FOL without identity.

The contents of sentences of $\mathcal{L}$ will be functions whose values are 1 and 0, individuated extensionally—that is, contents will be identified by the ordered pairs they contain. The question is, what are the domains of propositions, and how are they determined? The proposed answer is that for a given sentence ϕ of $\mathcal{L}$, the domain of the proposition expressed by ϕ is a partition of logical space into discrete cells, where the cardinality of the partition is determined by the number of atomic sentences occurring at least once in ϕ. Start with a single atomic sentence p_1 in $\mathcal{L}$. The proposition expressed by p_1 is the function mapping two points to $\{0, 1\}$, where the points represent the partition of logical space into two possibilities—the possibility that p_1 and the possibility that $\neg p$. The domain of the proposition expressed by p_1 is represented as the set of vertices of an instance of the binary hypercube graph Q_1:

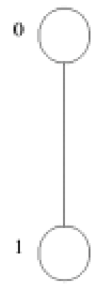

Figure 1

The top and bottom nodes represent the set of possible worlds supporting the falsity and truth of p_1, respectively. Given this domain, the content expressed by the atomic sentence p_1 is simply the function mapping the top node to 0 and the bottom node to 1. Suppose that the sentence ϕ is a truth-functional combination of atomic sentences p_1 and p_2, and nothing more. Then the domain of ϕ is the partition of logical space into four cells that represent the four possible combinations of truth-values of p_1 and p_2. We represent the domain of ϕ as the set of vertices of the Q_2 graph:

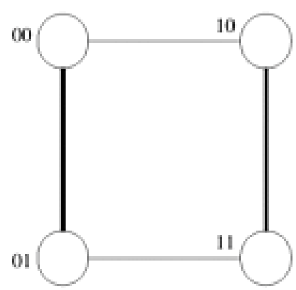

Figure 2

The content of ϕ is the function mapping the set of vertices of (the appropriate instance of) Q_2 into $\{0, 1\}$. It should be clear from the graph diagrams that the domains of p_1 and p_2 are simply projections of the domain of ϕ onto the first and second dimension of Q_2, respectively.[5]

Let ψ be a truth-functional combination of atomic sentences p_1, p_2, and p_3, and nothing more. Then the domain of ψ is a set of vertices of the binary hypercube Q_3:

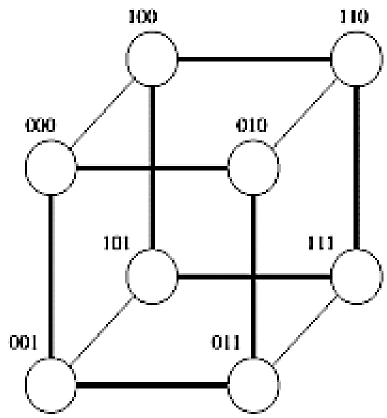

Figure 3

And if γ is a truth-functional combination of p_1—p_4, and nothing more, its domain is the set of vertices of an instance of Q_4—the tesseract:

[5]That is how it is possible to maintain that the domains (and, therefore, the *contents*) of p_1 and p_2 are distinct, even though they are both identified with an instance of Q_1. Outside of a higher-dimensional context, they would both be represented as the set of Q_1's vertices. But of course neither has a claim to being the privileged domain that is identifiable with *the* unique set of vertices of Q_1—the domains of p_1 and p_2 are both sets that are the sets of vertices of distinct instances of Q_1.

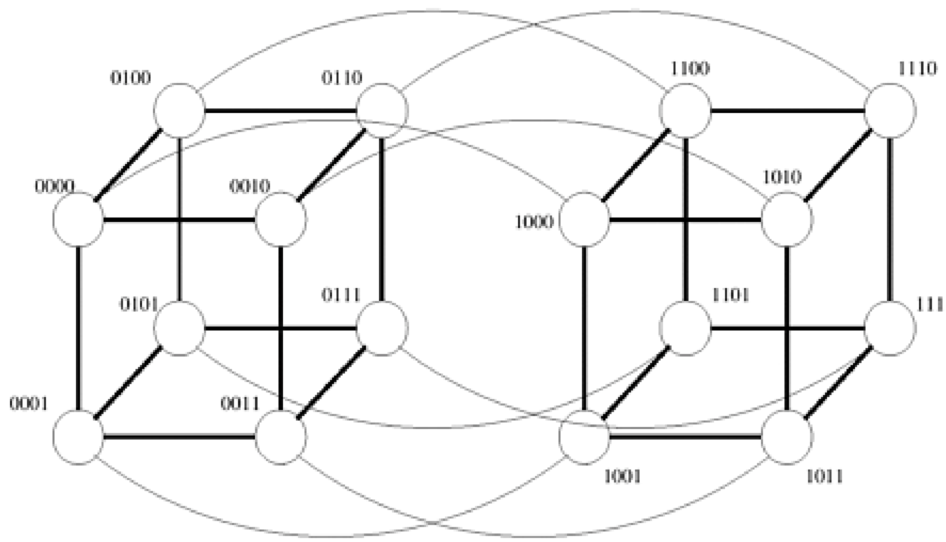

Figure 4

In general, for any k, the content of an $\mathcal{L}$-sentence containing at least one occurrence of exactly k distinct atomic sentences is a function whose domain is represented as a particular n-dimensional projection of some higher-dimensional space. If $\mathcal{L}$ has n atomic sentences, the total number of such functions is $\sum_{k=1}^{n} \binom{n}{k} 2^k$.

A note of caution: Although it can be a useful heuristic to think in terms of an analogy with ordinary truth tables, according to which the nodes of the graph Q_n correspond to the rows of a truth table with n propositional variables, the subtle disanalogy with truth table representations is also important. The nodes of the Q_i's represent points relative to which *fixed* atomic propositions and logical combinations thereof are assigned semantic values 1 or 0. Truth tables, on the other hand, do not depend on any choice of particular collections of atomic formulas. Rather, the "formulas" in the leftmost columns of the top row of a truth table are propositional metavariables. A particular domain with cardinality n of a proposition in the present sense is one of $\binom{m}{n}$ possibilities projected from Q_m. Contrast this with the fact that there is only one truth table with n rows (ignoring differences in the columns with logically complex "formulas" and possible differences in naming conventions for metavariables).

Here is how the familiar recursive assignment of possible world intensions to sentences of $\mathcal{L}$ would go. The recursive definition of semantic types and domains is:[6]

(1) SEMANTIC TYPES

a. e and t are semantic types.

b. If σ and τ are semantic types, then $\langle \sigma, \tau \rangle$ is a semantic type.

c. If σ is a semantic type, then $\langle s, \sigma \rangle$ is a semantic type.

d. Nothing else is a semantic type.

(2) SEMANTIC DOMAINS

a. D_e = the set of all individuals.

b. $D_t = \{0, 1\}$, the set of truth-values.

c. If σ and τ are semantic types, then $D_{\langle \sigma, \tau \rangle}$ is the set of all functions from D_σ to D_τ.

d. Intensions: If σ is a type, then $D_{\langle s, \sigma \rangle}$ is the set of all functions from W to D_σ.

An expression whose intension is in the domain $D_{\langle s, \sigma \rangle}$ is said to have an intension *of type* $\langle s, \sigma \rangle$. In $\mathcal{L}$, constants have intensions of type $\langle s, e \rangle$ (that is, constants denote functions from possible worlds to individuals in D_e); unary predicates have intensions of type $\langle s, \langle e, t \rangle \rangle$; binary predicates have intensions of type $\langle s, \langle e, \langle e, t \rangle \rangle \rangle$ sentences

[6] We are following the notation used in Von Fintel and Heim (2002). The reason for defining domains and types for sub-sentential constituents is to allow for the addition of quantifiers to $\mathcal{L}$ in the following section.

have intensions of type $\langle s, t\rangle$. The intensions of sentences of $\mathcal{L}$ are determined by the extensions of their parts relative to possible worlds:

(3) a. $\llbracket Pa_1...a_n \rrbracket_{\text{Ç}} = \lambda w_s.\ \llbracket P \rrbracket_{\text{Ç}}(w)(\llbracket a_1 \rrbracket_{\text{Ç}}(w))...(\llbracket a_n \rrbracket_{\text{Ç}}(w)) = 1.$

b. For $\mathcal{L}$-sentences ϕ and ψ:

i. $\llbracket \neg\phi \rrbracket_{\text{Ç}} = \lambda w_s.\ \llbracket \phi \rrbracket_{\text{Ç}}(w) = 0.$

ii. $\llbracket \phi \wedge \psi \rrbracket_{\text{Ç}} = \lambda w_s.\ \text{MIN}(\llbracket \phi \rrbracket_{\text{Ç}}(w), \llbracket \psi \rrbracket_{\text{Ç}}(w)).$

iii. $\llbracket \phi \vee \psi \rrbracket_{\text{Ç}} = \lambda w_s.\ \text{MAX}(\llbracket \phi \rrbracket_{\text{Ç}}(w), \llbracket \psi \rrbracket_{\text{Ç}}(w)).$[7]

Accordingly, the intensions of all sentences of $\mathcal{L}$ are functions defined over a fixed domain of possible worlds W. We now revise this recursive definition of intensions in light of the previous discussion so that they are functions from appropriate partitions of W to truth-values.

We begin by adding to our stock of primitive semantic types. Suppose that each of the atomic sentences of $\mathcal{L}$ are assigned a unique natural number. Then every set of atomic sentences is associated with a subset of $\mathbb{N}$. Certain subsets Γ of $\mathbb{N}$ induce a partition of W according to the atomic sentences encoded by members of Γ. Supposing that, say, $\Gamma = \{3, 19, 36\}$, the partition of W induced by Γ—denoted by W/Γ—has eight cells corresponding to all of the truth functional combinations of the three atomic sentences assigned respectively to 3, 19, and 36. So, one cell of W/Γ is the set of all members of W in which all three sentences are true; another cell will be the set of all members of W in which the first two are true and the third is false, and so on. Let $W^{\mathcal{L}}$ denote the set of all partitions of W induced by sets of natural numbers encoding the atomic sentences of $\mathcal{L}$. (Notice that if $\mathcal{L}$ contains n atomic sentences, then the cardinality of $W^{\mathcal{L}}$ is $\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1$.) We have the following

[7]We let $\llbracket \phi \rightarrow \psi \rrbracket_{\text{Ç}} := \llbracket \neg\phi \vee \psi \rrbracket_{\text{Ç}}$ and $\llbracket \phi \leftrightarrow \psi \rrbracket_{\text{Ç}} := \llbracket (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi) \rrbracket_{\text{Ç}}$.

expanded inventory of primitive types:

(4) Semantic Types*

a. e and t are semantic types.

b. If σ and τ are semantic types, then $\langle \sigma, \tau \rangle$ is a semantic type.

c. If σ is a semantic type, then $\langle s, \sigma \rangle$ is a semantic type.

d. If Γ is a set of natural numbers encoding atomic sentences of $\mathcal{L}$, then $\langle s_\Gamma, t \rangle$ is a semantic type.

e. Nothing else is a semantic type.

The new semantic types, in turn, expand our inventory of semantic domains:

(5) Semantic Domains*

a. $D_e = D$, the set of all individuals.

b. $D_t = \{0, 1\}$, the set of truth-values.

c. If σ and τ are semantic types, then $D_{\langle \sigma, \tau \rangle}$ is the set of all functions from D_σ to D_τ.

d. Intensions: If σ is a type, then $D_{\langle s, \sigma \rangle}$ is the set of all functions from W to D_σ.

e. Contents: If Γ is a set of natural numbers encoding atomic sentences of $\mathcal{L}$, $D_{\langle s_\Gamma, t \rangle}$ is the set of all functions from W/Γ to $\{0, 1\}$.

It is already possible to see that contents, so-defined, are more fine-grained than possible world intensions (functions of type $\langle s, t \rangle$). If $\mathcal{L}$ has n atomic sentences, then the total number of contents that are expressible in the language would be

$\sum_{k=1}^{n} \binom{n}{k} 2^n$—by contrast, there would only be 2^n intensions expressible in $\mathcal{L}$.[8]

(6) (Definition) If w is a member of W/Γ, and $\Gamma' \subseteq \Gamma$, then $\text{proj}_{\Gamma'}(w)$ is the unique member w' of W/Γ' such that $w \subseteq w'$.

The $\text{proj}_X()$ functions are functions that project cells in a "higher-dimensional" partition of W onto cells in a partition determined by a subset of those dimensions. For illustration, picturing the nodes of the hypercube graph Q_3 as the eight cells of a partition of W by three atomic formulae, the function $\text{proj}_\Gamma()$ is one of the projections of Q_3 onto Q_2 or Q_1, the nodes of which represent cells of a lower-dimensional partition of W. We now have the machinery necessary for a recursive assignment of contents to $\mathcal{L}$-sentences ϕ, which we will denote with $[\![\phi]\!]_*$:

(7) a. For p_i the i^{th} atomic sentence of $\mathcal{L}$: $[\![p_i]\!]_* = \lambda w_{s_{\{i\}}}.\ \forall w' \in w$: $[\![p_i]\!]_{\c{C}}(w') = 1$.

b. For $\mathcal{L}$-sentences ϕ and ψ of type $\langle s_\Phi, t\rangle$ and $\langle s_\Psi, t\rangle$ respectively:

i. $[\![\neg\phi]\!]_* = \lambda w_{s_\Phi}.\ [\![\phi]\!]_*(w) = 0$.

ii. $[\![\phi \wedge \psi]\!]_* = \lambda w_{s_{\Phi\cup\Psi}}.\ \text{MIN}([\![\phi]\!]_*(\text{proj}_\Phi(w)), [\![\psi]\!]_*(\text{proj}_\Psi(w)))$.

iii. $[\![\phi \vee \psi]\!]_* = \lambda w_{s_{\Phi\cup\Psi}}.\ \text{MAX}([\![\phi]\!]_*(\text{proj}_\Phi(w)), [\![\psi]\!]_*(\text{proj}_\Psi(w)))$.[9]

Just where a particular sentence's content lives in the hierarchy of types is determined by the types of its parts. This marks an important distinction of the present account. Whereas according to theories identifying contents with possible world intensions, the contents of sentences were all of the same type $\langle s, t\rangle$, a peculiarity of the resulting contents assigned to sentences is that they are functions defined over shifty domains.

[8] This is analogous to the difference between the two different ways of individuating the possible states of a machine with n distinct binary components.

[9] Similarly, we let $[\![\phi \rightarrow \psi]\!]_* := [\![\neg\phi \vee \psi]\!]_*$ and $[\![\phi \leftrightarrow \psi]\!]_* := [\![(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)]\!]_*$

A given $\mathcal{L}$-sentence ϕ, according to the present account, can be seen to have three distinct types: First, the extension of the ϕ is simply a truth value 0 or 1; second, the intension of a sentence is that function of type $\langle s, t \rangle$ mapping W into $\{0, 1\}$, third, the content of ϕ is that function of type $\langle s_\Gamma, t \rangle$ for some $\Gamma \subseteq \mathbb{N}$, determined by the ϕ's parts, mapping W/Γ into $\{0, 1\}$. We conclude the current section with some observations and clarification remarks on the construction in (7).

The analogy between contents with the states of the representational system described in the previous section should be evident—both can be modeled as functions mapping the nodes of binary hypercubes of in some finite dimension into truth values. We noted that the states of the representational system can be individuated, in part, by the particular set of modules that are active. So the nodes of a hypercube in n dimensions represent the 2^n possible configurations of each of its binary component modules. Similarly, the contents of sentences can be individuated, in part, by the set of atomic formulas occurring in them. The nodes of a hypercube in n dimensions represent the 2^n cells of the partition of logical space induced by the set of "active" atomic formulas.

Does it follow that the characterization of content in (7) involves conflating form with content?[10] After all, the contents of sentences are being individuated, in part, by the set of atomic formulas occurring in them. It looks as though some of the explicit syntax of $\mathcal{L}$ is contaminating contents, something like the way in which syntactic structure is mirrored by Russellian structured propositions. But this is incorrect. It is clear in (7) that the ordinary intensions of atomic formulae determine the contents of atomic formulae, which are defined to be functions defined on various partitions of

[10]Recall that this is one problem with structuralist accounts of propositions—cf. Chapter 2, sect. 2.

W—the cells of which are comprised of ordinary possible world intensions (see (7a)). There is nothing special about the atomic sentences *themselves* that plays a role in the construction of contents. At bottom, it is the ordinary intensions of atomic formulae that determine the relevant partitions of W featured in the recursive clauses of (7). (Another way to appreciate the same point is to consider a language $\mathcal{L}'$ that is identical to $\mathcal{L}$ except that its atomic formulae are notationally different from those of $\mathcal{L}$. Let f be a one-to-one correspondence between the atomic formulae of $\mathcal{L}'$ and those of $\mathcal{L}$ such that for all p in Dom(f): $\llbracket p \rrbracket_{\text{Ç}} = \llbracket f(p) \rrbracket_{\text{Ç}}$. Notice that, from the contents assigned to complex sentences of $\mathcal{L}'$ by (7), we can determine the contents of complex sentences in $\mathcal{L}$ as follows: for ϕ a complex sentence in $\mathcal{L}$, $\llbracket \phi \rrbracket_* = \llbracket f^{-1}(\phi) \rrbracket_*$ where $f^{-1}(\phi)$ is that sentence of $\mathcal{L}'$ formed by replacing all atomic sentences p occurring in ϕ with $f^{-1}(p)$.)

3.3 Solutions to Problems

Propositional contents as defined in the previous section have a number of advantages over possible worlds propositions. We start with the most pressing of the problems discussed in the previous chapters.

3.3.1 Nonidentity of Logical Equivalents

According to the familiar relational accounts, a propositional attitude report is true just in case the subject of the report stands in a particular binary relation to the proposition expressed by the complement clause of the report. It follows that any two clauses expressing the same proposition are interchangeable salva veritate in the complements of attitude reports. Perhaps the most damaging consequence of the intension/content identification is the fact that, together with the relational account of attitude reports, it predicts that logically equivalent clauses are identical and, therefore, interchangeable salva veritate in the complements of attitude reports. The previous chapter reviewed some of the ways for the defender of coarse propositions to deal with this unattractive result (idealization, fragmentation, diagonalization, appeals to pragmatics), none of which were ultimately very satisfying. The solution to the problem of interchangeability calls for fine-grained propositions of some sort or other.

To show that logically equivalent contents, as defined in (7) above, need not be interchangeable in propositional attitude reports, it suffices to show that there are logically equivalent propositions that are nevertheless not identical. Consider the logically equivalent sentences p_i and $p_i \vee (p_k \wedge \neg p_k)$, where p_i and p_k are atomic $\mathcal{L}$-sentences assigned to i and k respectively. According to (7), the content of p_i is that function of type $\langle s_{\{i\}}, t\rangle$ that maps partition cells of W to the value 1 just in case every element of that cell is a member of the ordinary intension of p_i. By contrast, the content assigned to $p_i \vee (p_k \wedge \neg p_k)$ is that function of type $\langle s_{\{i,k\}}, t\rangle$ that maps partition cells w of W to 1 depending on the values of $[\![p_i]\!]_*(\text{proj}_{\{i\}}(w))$ and $[\![(p_k \wedge \neg p_k)]\!]_*(\text{proj}_{\{k\}}(w))$. These contents are distinct because functions are individuated extensionally by their members. That being the case, it is possible for an agent to stand in a propositional attitude relation to one and not the other. This is how the problem of interchangeability of necessary equivalents in attitude reports is resolved. Even though the propositions are necessarily equivalent (they have the same possible worlds intensions), they are defined on extensionally distinct domains.

Information and Logical Necessity

The following two sentences are about different things:

(8) a. Either it will rain today, or it will not.

b. Either the stock market will crash today, or it will not.

The former is concerned with whether or not it will rain today and the latter is concerned with whether or not the stock market will crash today. Accordingly, since they concern different subjects, an agent might believe or assert (8a) without thereby believing or asserting (8b).[11] For example, a child having no concept of the stock market might have logically true beliefs about today's rain without having *any* beliefs about the stock market, let alone logically true ones. But according to coarse-grained accounts of propositional content, these intuitions about the distinction between contents expressed by (8a) and (8b) must be disregarded or explained away—the content of (8a) and (8b) is simply the necessary proposition, the constant function mapping worlds to the value 1.

But since, according to (7), logically equivalent sentences need not express the same contents, we are not automatically forced to say that all logically necessary contents are the same. In particular, supposing that (8a) and (8b) are translated as disjunctions of atomic $\mathcal{L}$-sentences, their contents would be distinct from one another because they are functions defined on distinct domains: (8a) maps the two partition cells of W determined by 'It will rain today' to the value 1, while (8b) maps the two partition cells of W determined by 'The market will crash today' to the value 1. So, not only are there logically equivalent sentences that can express distinct contents,

[11] We are bracketing the possibility of idiosyncratic psychological tendencies.

there are, in particular, pairs of logically true sentences that express distinct contents. It follows that at epistemic agent might bear a cognitive attitude towards one necessary truth without thereby bearing that attitude towards *all* necessary truths. This shows that at least two versions of the logical omniscience problem plaguing unstructured coarse-grained contents are overcome by our present characterization of unstructured contents.

It might be objected that there is a sense in which (8a) and (8b) are about the same thing and, therefore, express the same content—they are about everything since they express logical truths and logical truths do not have a subject matter. Someone who knows only the meanings of truth-functional disjunction and negation would know that (8a) and (8b) are both true, regardless of the contents of their constituents. Logical truths like (8a) and (8b) are topic neutral; so any theory allocating distinct contents to distinct logically true sentences must be incorrect.

While there is certainly some sense in which pairs of logical truths are not about distinct things (i.e. the sense in which their truth conditions are neutral with respect to the contents of their constituents), this is not the sense of *aboutness* relevant in evaluating (7). There is, after all, a similar sense in which arithmetic truths are about everything, at least everything that can be numbered, counted, measured, etc., but it does not follow that arithmetic is not about the natural numbers. The problem with this objection is the presupposition that understanding the conditions under which a sentence is true is not only necessary but *sufficient* to grasp the content expressed by a sentence. But anyone who accepted the latter—that understanding truth conditions is sufficient for understanding content—would certainly be someone already committed to a coarse-grained account of propositional content according to which all necessarily true sentences express the same content, and this is just one of the consequences of coarse-grained contents that motivates a finer characterization of propositional contents.

3.3.2 Logical Consequence

Recall that unstructured content forces certain propositional attitude predicates—those that distribute over conjunction—to be closed under logical consequence. [12] But this paper has been a defense of a version of unstructured content; does this argument show that (distributive) propositional attitude predicates must be closed under logical consequence after all? No. According to (7) even if $\phi \models \psi$, it does not follow that $[\![\phi]\!]_*^{-1}[\{1\}] \subseteq [\![\psi]\!]_*^{-1}[\{1\}]$ (that is, it does not follow that the set of arguments to $[\![\phi]\!]_*$ that get mapped to 1 is a subset of the set of arguments to $[\![\psi]\!]_*$ that get mapped to 1). So the premise of the previous argument identifying the content of ϕ with the content of $\phi \wedge \psi$ is false under the present characterization of content. So distributive propositional attitude predicates need not be closed under logical consequence. Consequently, this argument aimed at theories of unstructured contents does not override the foregoing observations: necessarily equivalent sentences—and, in particular, sentences expressing logical truths—need not be interchangeable in the complement clauses of propositional attitude reports.

[12] Here is the argument: Let ϕ and ψ be $\mathcal{L}$-sentences such that the set of worlds supporting the truth of ϕ is a subset of the set of worlds supporting the truth of ψ. Let P be a propositional attitude predicate that distributes over conjunction and suppose that $P(\phi)$. Since the possible worlds content of $\phi \wedge \psi$ is the intersection of that of ϕ and ψ, the possible worlds content of $\phi \wedge \psi$ is just that of ϕ itself. It follows that $P(\phi \wedge \psi)$. Since P distributes over conjunction, it follows that $P(\psi)$, showing that P is closed under logical consequence.

3.3.3 Revenge Cases?

Even though the problems that face coarse-grained unstructured intensions do not apply in full generality to our present characterization of content, it is conceivable that there are special instances of these problems that still apply. What we have shown so far is that contents defined on appropriate partitions of the domain W allow for fine-grained distinctions between logically equivalent contents. This is consistent, however, with the existence of distinct $\mathcal{L}$-sentences the contents of which are defined over the same partition of W. This is to be expected—otherwise all syntactically distinct sentences would express distinct contents. But, because of this, the threat of conflating distinct contents could conceivably reemerge. If $[\![\phi]\!]_*$ and $[\![\psi]\!]_*$ are defined over the same partition of W, all three of the following problems do reemerge: (i) if ϕ and ψ are logically equivalent sentences, they are identical in content, and are, therefore, interchangeable in the complements of propositional attitude reports; (ii) if they're both logically necessary sentences, they express the same contents; (iii) the argument from logical consequence through as normal—whenever $[\![\phi]\!]_*^{-1}[\{1\}] \subseteq [\![\psi]\!]_*^{-1}[\{1\}]$, we get, according to (7bii), that the content of ϕ is identical to the content of $\phi \wedge \psi$—so propositional attitude predicates that distribute over conjunction are closed under logical consequence (limited to sentences defined over a given partition of W).

It is not hard to see that, in fact, these are *favorable* predictions of our present characterization of contents. To see why, consider an example of the prediction in (i), in which the following sentences are predicted to express the same content:

(9) a. Grass is green and Snow is white.

b. Snow is white and grass is green.

These sentences express the same content—merely reordering conjuncts does not intuitively result in a change of content.[13] In fact, it seems that any arbitrary logically equivalent formulation of (9a) and (9b) that contains exactly the atomic sentences 'Grass is green' and 'Snow is white' (and no others) will amount to a rewording of just the same content.

Is there any reason for holding that logically equivalent but syntactically distinct rewordings of (9a) or (9b) could express a distinct content? First, if such a pair *could* express distinct contents, it would not be in virtue of having distinct truth conditions—we are only considering pairs of logically equivalent sentences. Second, it is difficult to maintain that any pair of such sentences could be intuitively *about* different things. If two sentences contain exactly the same atomic sentences as constituents, and they are logically equivalent, then any distinction in their contents must be traceable to their logical expressions. But (9a) and (9b) are about grass and snow, not *conjunction*—and it seems that the same could be said for any other logical permutation of them. Third, if it is possible for an agent to believe the content expressed by (9a) (or (9b)) without thereby believing the content expressed by some logically equivalent formulation containing just the atomic sentences occurring in (9a) (and no more), then the relational account of belief dictates that different contents are being expressed. But it is not much of a stretch to suggest that (9a) and (9b) are themselves mutually interchangeable salva veritate in the complements of belief ascriptions. Moreover, since it is implausible that introducing more logical complexity to a sentence could play a role in altering content if no new atomic formulas are introduced (and truth conditions are preserved), we can safely generalize the claim

[13] (9a) and (9b) are predicted to have identical contents only if the conjunctions are tenseless. That is, e.g., (9a) does not mean that grass is green *and then* snow is white.

about interchangeability within complement clauses of belief reports: if ϕ and ψ are logically equivalent and are composed of the atomic formulas, any agent believes that ϕ just in case it believes that ψ.

In fact, apart from whether or not our characterization of content predicts this general interchangeability of logical equivalents relative to a particular partition of W, it can be independently established by induction on the complexity of $\mathcal{L}$-sentences as follows: Let S be any epistemic agent and suppose that the $\mathcal{L}$-sentence A is composed of the two atomic $\mathcal{L}$-sentences p_i and p_k. The base case—S believes that A—is given. Suppose that B is any logically equivalent formulation of A composed of only p_i and p_k. By inductive hypothesis, S believes that B. Now suppose that B' is logically equivalent to B, that B' is composed of atomic sentences p_i and p_k, and that the logical complexity of B' is exactly one degree greater than that of B. Then, since one degree of logical complexity does not change content if we are holding truth conditions and component atomic formulas fixed, it follows that S believes B'.[14]

So, why doesn't an analogous inductive argument establish that *any* logically equivalent $\mathcal{L}$-sentences would be mutually interchangeable in the complements of belief ascriptions? The reason this argument does not generalize to any logically equivalent $\mathcal{L}$-sentences is simply that the inductive step would become false. The support for the inductive step in the present argument is that adding degrees of logical complexity to a sentence while holding everything else fixed amounts to a trivial rewording of that sentence. But in a generalized version of the inductive step, everything else is *not* held fixed—there is no prohibition against introducing new atomic sentences in addition to the introduction of more logical complexity. But, in contrast to merely

[14]Recall that this was proposed as a plausible individuation condition for propositions in chapter 1, section 3.

introducing new degrees of logical complexity, examples like (8a)/(8b) above illustrate that substituting the atomic sentences composing a complex sentence is sufficient to induce a change in content. So, the cases in which our definition of content predicts interchangeability in belief reports—cases in which truth-conditions and atomic constituents are held fixed—seem to be harmless predictions of the theory. In fact, they are just the cases we would expect to fall out of a nontrivial predictive theory of propositional content.

3.3.4 A Solution to the Frege-Soames Puzzle

We show here how the present construction of contents can be used to solve the Frege-Soames puzzle, which undermines worlds-based theories of propositional content at least as much as logical omniscience phenomena.

Quantification

The solution to be given also serves to illustrate the way to extend the present semantics, which only applies to the quantifier-free language $\mathcal{L}$, to the language $\mathcal{L}^*$ ($\mathcal{L}$ with quantifiers). The semantic denotations of the connectives given in the previous section lets connectives operate on the contents expressed by component sentences. The content of conjunction, for example, is a function operating on the contents of the component conjuncts, which are functions from partitions of W to $\{1, 0\}$. In this way, we are able to ensure that the semantic types of connectives are sensitive to the types of their component sentences. Since the semantic types of sentences are sensitive to their parts, it should not be surprising that the semantic types of connectives are required to shift depending on the components being connected. The approach of letting connectives operate on the contents of components suggests a way to extend the present semantics to assign denotations to the quantifiers.

If quantifiers are to operate on contents of open sentences, it is necessary to first explain what the *content* of an open sentence is supposed to be. So far, the only things that have been assigned contents (functions from partitions of W to extensions) are closed sentences and connectives. Let $P(x)$ be an open atomic sentence of $\mathcal{L}$. Rather than letting $\exists x P(x)$ and $\forall x P(x)$ be built up by prefixing $\exists x$ and $\forall x$, respectively, to $P(x)$, we will assume that the bare quantifiers $\exists$ and $\forall$ are prefixed to the predicate abstract $xP(x)$. This means that the bare quantifiers and $xP(x)$ need to be assigned semantic denotations such that one of them takes the other as an argument.[15]

How should we think about the content of the predicate abstract $xP(x)$? If we think of its ordinary extension as that of an open sentence, its ordinary extension would be a function that maps individuals to truth values—something of type $\langle e, t \rangle$. This means that its ordinary intention would be a function from possible worlds to functions of type $\langle e, t \rangle$. This suggests defining the content of $xP(x)$ as a function mapping cells of a partition of W to ordinary extensions. But what partition of W is the content of $xP(x)$ defined on in the first place? For clarity, lets assume that $P(x)$ is $smokes(x)$. We can work backwards from the ordinary extensions of $x\ [smokes(x)]$, relative to possible worlds. In each possible world there is a set of individuals who are smokers at that world; in some possible worlds everyone is a smoker, in others nobody is a smoker. Let's suppose that Jane and Bob are the only smokers at w. There are other possible worlds that are like w in this respect. These worlds will

[15] We assume the English-like phrase structure of quantified first-order sentences given in Heim and Kratzer (1998, ch 7.4). It is English-like insofar as it allows for a distinction between determiner phrases—e.g. 'every man'—and bare determiners—e.g. 'every'.

constitute one partition cell of W. In this way, the predicate abstract $smokes(x)$ can be said to induce a partition of W, where the equivalence relation corresponding to the partition is the *has-the-same-smokers-as* relation. In general, if P is the i^{th} atomic predicate of $\mathcal{L}$, we will let $w_{s_{\{i\}}}$ denote an arbitrary member of the partition of W induced by $xP(x)$.

Given this way of identifying partitions of W based on predicate abstracts, we can now define the content of the predicate abstract component of $\exists xP(x)$.

(10) For P_i the i^{th} atomic predicate of $\mathcal{L}$: $[\![x\ \ P_i(x)]\!]_* = \lambda w_{s_{\{i\}}}.\lambda x_e.\forall w' \in w$: $([\![wP_i]\!]_{\c{C}}(w'))(x) = 1$.

The content of $\exists xP(x)$, as planned, is defined as a function mapping cells of an appropriate partition of W onto extensions.

This generalizes beyond atomic predicate abstracts in the obvious way. For example, the predicate abstract $x[smokes(x) \wedge drinks(x)]$ will induce a partition of W given by the equivalence relation *has-the-same-individuals-who-both-drink-and-smoke*. Under the partition of W induced by $x[smokes(x) \wedge drinks(x)]$, two worlds w and w' will be in the same partition cell just in case every individual that both drinks and smokes in one is also an individual who drinks and smokes in the other. We have the corresponding generalization of (15):

(11) $[\![x[\phi(x)]]\!]_* = \lambda w_{s_\Phi}.\lambda x_e.\forall w' \in w$: $([\![wP_i]\!]_{\c{C}}(w'))(x) = 1$ (where s_Φ denotes the partition of W determined by the predicate abstract $x[\phi(x)]$)

Keeping with the strategy of letting the semantic denotations of connectives operate on contents of components, we can now provide the recursive clauses for quantified sentences, which are predicate abstracts with the bare quantifiers $\exists$ and $\forall$ prefixed:

(12) When affixed to a predicate abstract $x[\phi(x)]$ of type $\langle s_\Phi, \langle e, t\rangle\rangle$:

a. $[\![\exists]\!]_* = \lambda f_{\langle s_\Phi, \langle e,t\rangle\rangle}.\lambda w_{s_{\{i\}}}.\ \{x : f(w) = 1\}$ is non-empty.

b. $[\![\forall]\!]_* = \lambda f_{\langle s_\Phi, \langle e,t\rangle\rangle}.\lambda w_{s_{\{i\}}}.\ \{x : f(w) = 1\}$ is the domain of all individuals.

It is not hard to see that the contents of quantified sentences have the types corresponding to the types of their embedded predicate abstracts: $[\![\exists x[\phi(x)]]\!]_* =$

$= [\![\exists]\!]_*([\![x[\phi(x)]]\!]_*)$

$= [\lambda f_{\langle s_\Phi, \langle e,t\rangle\rangle}.\lambda w_{s_{\{i\}}}.\{x : f(w) = 1\}$ is non-empty.$]([\![x[\phi(x)]]\!]_*)$

$= \lambda w_{s_{\{i\}}}.\{x : [\![x[\phi(x)]]\!]_*(w) = 1\}$ is non-empty.

Hence, the content $[\![\exists x[\phi(x)]]\!]_*$ has the type of $[\![x[\phi(x)]]\!]_*$, which is simply $\langle s_\Phi, t\rangle$

Application

If proper names denote their referents rigidly, this raises semantic puzzles for everyone—at least everyone who accepts that propositional attitude predicates relate agents to propositional contents. Defenders of Russellian structured propositions no less than defenders of unstructured propositions have to face Frege's puzzle. Consider the following principle of compositionality:

(13) (Comp) If S_1 and S_2 are non-intensional sentences or formulas with the same grammatical structure, which differ only in the substitution of constituents with the same semantic contents (relative to their respective contexts and assignments), then the semantic contents of S_1 and S_2 will be the same (relative to those contexts and assignments).

For example, if the names 'Hesperus' and 'Phosphorus' rigidly co-refer so that they can be interchanged without inducing a change in content, the following express the same content:

(14) a. Hesperus is the morning star.

b. Phosphorus is the morning star.

Accordingly, (10a) and (10b) should be interchangeable salva veritate in belief contexts. This conflicts with the intuition that (11a) is true, but (11b) is false.

(15) a. The ancients believed that Hesperus is the morning star.

b. The ancients believed that Phosphorus is the morning star.

So, anybody accepting the relational theory of belief (and other attitudes) together with direct reference, and the correlative commitment to the preservation of content through substitution of co-referring names, faces this version of Frege's puzzle.

The aim here is not to solve Frege's puzzle in full generality but to show that the devastating strengthening of Frege's puzzle in **?**, which undermines all theories according to which propositions are unstructured sets of truth-supporting circumstances, does not touch *our* version of unstructured contents. Here is a quick review of we have been calling the Frege-Soames puzzle. We start with some commitments that are shared by defenders of unstructured contents:

(16) a. The semantic content of a sentence (relative to a context) is the collection of circumstances supporting its truth (as used in the context).

b. Propositional attitude sentences report relations to the semantic contents of their complements—i.e. an individual i satisfies 'x vs that S' (relative to a context C) iff i bears an appropriate binary relation R to the semantic content of S (relative to C).

c. Many propositional attitude verbs, including 'say', 'assert', 'believe', 'know', and 'prove' distribute over conjunction.

d. Names, indexicals and variables are directly referential.

Then, assuming that (17a) is true, we can derive the absurd conclusion (17d) with the use of (16a)—(16d):

(17) a. The ancients believed the 'Hesperus' referred to Hesperus and 'Phospohorus' referred to Phosphorus. (Assumed)

b. The ancients believed that 'Hesperus' referred to Phosphorus and 'Phosphorus' referred to Phosphorus. (By (16a), (16b), (16d) and (Comp))

c. The ancients believed that 'Hesperus' referred to Phosphorus and 'Phosphorus' referred to Phosphorus, and there exists an x such that 'Hesperus' referred to x and 'Phosphorus' referred to x."

d. The ancients believed that there exists an x such that 'Hesperus' referred to x and 'Phosphorus' referred to x. (By (16c))

We are assuming that the complements of each of the belief ascriptions in (17a)—(17d) can be expressed by appropriate $\mathcal{L}$-sentences. If the contents of a given $\mathcal{L}$-sentence were identical with the set of possible worlds in which it is true, then, indeed, the assumed truth (17a) would lead to the intuitive falsity (17d).

According to the semantic denotations of quantified sentences above, the inference from (17b) would be (17c) invalid. If the truth-supporting circumstances were worlds, possible or impossible, the justification for inferring (17c) from (17b) comes from a principle about existential generalization:

> **(Gen)** For all worlds w, a sentence ϕ is true at w only if $\exists x\phi$ is true at w.

Indeed, (Gen) seems to be constitutive of existential quantification. However, if sentential contents are defined as functions mapping partitions of W to truth values, the principle (Gen) would have to be reworded so as to quantify over partition cells instead of quantifying over worlds directly.

> **(Gen*)** Given a partition p of W, a sentence ϕ is true relative to a member of p only if $\exists x\phi$ is true relative to a member of p.

Not only does the principle (Gen*) lack the intuitive appeal that motivates the principle (Gen), but according to the semantics for existential quantification given above, (Gen*) is false. Consider, for example, the atomic sentence 'Bob runs' and the corresponding atomic predicate abstract x [x runs]. The content of the 'Bob runs' is a function defined on the partition of W consisting of exactly two cells—one cell containing the possible worlds in which Bob runs and one cell containing the worlds in which Bob does not run. The content of x [x runs] is a function defined on the partition of W determined by the *has-the-same-runners* relation—the number of cells of which is well beyond two (if there are, say, n possible runners, the number of cells will turn out to be $\sum_{k=0}^{n} \binom{n}{k} = 2^n$). Since the domain of $[\![\exists x[x \text{ runs}]]\!]_*$ is identical to that of $[\![x[x \text{ runs}]]\!]_*$, we know that the truth of 'Bob runs' relative to a partition cell p does not entail the truth of $\exists x[x$ runs] relative to p—so (Gen*) turns out to be false.

So a key inference in the argument (17a)—(17d) is invalid if sentential contents are defined as being functions from appropriate partition cells to truth values. Other ways around the Frege-Soames puzzle either leave analogous problems unsolved, or are ad hoc denials of background assumptions like (Gen). For instance, if one rejects (16d) by denying that proper names are rigidly denoting, the argument (17a)—(17d) can be reformulated using only rigidly denoting indexical expressions.[16] If one rejects (16c) by claiming that belief fails to distribute over conjunction, then the argument

[16]See **?**, section 4.

(17a)—(17d) can be reformulated using any other propositional attitude predicate that *does* distribute over conjunction (unless one is prepared to deny distributivity over conjunction for *all* propositional attitude predicates). Moreover, the argument (17a)—(17c) is sufficiently devastating independent of any assumptions regarding distribution over conjunction.

It should be noted that defenders of impossible worlds propositions are free to reject (Gen), so that relative to some worlds a sentence can be true without the corresponding generalization being true.[17] There are at least two problems with this sort of ad hoc denial of (Gen). First, if, as argued in the previous chapter, impossible worlds are supposed to model epistemic possibilities instead of arbitrary collections of sentences, it is doubtful that such worlds should be counted among anyone's epistemic possibilities. Second, it seems that there should be some independent reason for rejecting (Gen) besides just its ineliminable role in the Frege-Soames puzzle. The semantic framework developed in this chapter predicts this from more basic principles about the individuation of propositional contents across sentences.

3.4 Open Questions

There are a few questions that are left open here for future work. For example, if propositions really are individuated in part on the basis of the ordinary intensions of atomic sentences, this raises the question of how the atomic sentences of a language are to be circumscribed. The construction of contents based on appropriate partitions of W presupposes something like the early Wittgensteinian atomistic thesis that all the sentences of a language are composed of logically independent constituent sentences.

[17]See, for example, Priest (1987).

In fact, Wittgenstein himself grew pessimistic about the possibility of specifying the atomic constituents of a language:

> I [believed that] the elementary propositions could be specified at a later date. Only in recent years have I broken away from that mistake. At the time I wrote in a manuscript of my book, "The answers to philosophical questions must never be surprising. In philosophy you cannot discover anything." I myself, however, had not clearly enough understood this and offended against it.[18]

Also, the color-exclusion problem and the possibility of non-logical necessary truths, could prove to be problematic for the ideas developed here. If there *are* such relations as necessary entailment or exclusion at the atomic level of a language (assuming there is some appropriate circumscription of the atomic sentences in the first place), then the granularity of contents defined here would seem to be too coarse from the start. Recent work on the color exclusion problem aimed at rectifying a broadly Tractatarian project could be useful in this regard—even if the philosophical aims are different from those here.

We have also been presupposing that there is a genuine distinction between the logical and non-logical expressions of a language such that the former are in some sense *content-neutral* expressions. This presupposition is what supports the claim that holding truth conditions and the atomic constituents of a sentence fixed, differences in logical syntax do not induce a change in content. But, the precise connection between, on the one hand, the supposed content neutrality of logical expressions and, on the other hand, the fact that such differences in logical syntax preserve content would need to be made more explicit in order to be a satisfying theoretical explanation of the intuitive data. One way of making this connection more explicit would be to

[18] Waismann (1979, 182)

survey various proposed criteria for demarcating the logical constants of a language for the sake of drawing out possible predictions. Of course, any explicit predictions of the insufficiency of distinctions in logical syntax as such for generating distinctions in propositional content would help to ground the ideas developed here in other, independently motivated, research projects.

Chapter 4

PARITY OF CLOSURE CONDITIONS

4.0

Consider the following pair of candidate principles about rational belief and knowledge:

> **($\mathbf{MP}_B$)** For any rational agent S, if S believes that A, and S believes that A entails B, then S believes that B.
>
> **($\mathbf{MP}_K$)** For any rational agent S, if S knows that A, and S knows that A entails B, then S knows B.

It is at least plausible that (MP_B) is a normative ideal governing rational belief. It is at least more plausible than the corresponding principle for knowledge (MP_K). After all, assuming knowledge entails some degree of justification, S's justification for believing the premises (that is, for believing that A and that A entails B), may only minimally qualify for S's having knowledge. If that were the case, an inferred conclusion depending on both premises would not be sufficiently justified to count as an item of knowledge.

Second, consider the following candidate principles about rational belief and knowledge:

> **(Con$_B$)** For any rational agent S, if S believes that A, and S believes that B, then S believes that A *and* B.
>
> **(Con$_K$)** For any rational agent S, if S knows that A, and S knows that B, then S knows that A *and* B.

Again, it seems possible to hold that (Con$_B$) is a normative ideal governing rational belief while denying (Con$_K$). An agent's knowledge set need not be closed under conjunction introduction unless the sort of justification required for knowledge is 100% indefeasible.

What do these pairs have in common? They are candidate closure conditions for rational belief and knowledge in which the belief versions are, at least *prima facie*, much more plausible than the knowledge versions. It would, therefore, be a surprising result if it turned out that, for any of these pairs, it is impossible to validate the belief principle without thereby validating the corresponding knowledge principle. Unfortunately, this is precisely what is predicted by the standard Kripke-style modal treatments of knowledge and belief, even if we generalize the semantics to include impossible worlds.

4.1 Some Modal Semantics

Let $\mathcal{L}$ be a language generated from countably many atomic sentences p_i together with the usual logical constants (and parentheses) via the usual recursive clauses. Let $\mathcal{L}_{B,K}$ be a language identical to $\mathcal{L}$ except that its alphabet includes the sentential operator B_S (standing for "agent S believes that...") and K_S (standing for "agent S knows that ..."), where well-formed formulas in the extended language $\mathcal{L}_{B,K}$ are generated in the obvious way. So $\mathcal{L}_{B,K}$ is a basic propositional language enriched with the sentential operators B_S and K_S.

The semantics for $\mathcal{L}_{B,K}$ is more or less identical to the usual semantics for the modal logic of necessity and possibility: A *frame* $\mathcal{F}$ is a triple (W, R_B, R_K) where W is a set of possible worlds and where R_B and R_K are both subsets of $W \times W$. If either R_B or R_K has a certain formal property, say R_B is symmetric, we say that $\mathcal{F}$ is an *R_B-symmetric frame*, and similarly for other properties of R_B and R_K. A *model* $\mathcal{M}$ over a frame $\mathcal{F}$ ($= (W, R_B, R_K)$) is a pair $(\mathcal{F}, \phi)$ in which ϕ, the *assignment function*, is a function from atomic sentences into $\mathcal{P}(W)$ (the powerset of $\mathcal{M}$'s *domain*). Truth of formulas in $\mathcal{L}_B$ is defined relative to model-world pairs. Let $\mathcal{M}$ be a model over $\mathcal{F} = (W, R_B, R_K)$ with assignment function ϕ and let $w \in W$. Then, for an $\mathcal{L}_{B,K}$-formula α, $\mathcal{M},w \models \alpha$ (i.e. α is true in model $\mathcal{M}$ at world w) is defined recursively as follows:

For an atomic sentence α: $\mathcal{M},w \models \alpha \Leftrightarrow w \in \phi(\alpha)$

For $\alpha = \neg\psi$: $\mathcal{M},w \models \alpha \Leftrightarrow$ not $\mathcal{M},w \models \psi$

For $\alpha = (\phi \wedge \gamma)$, $(\phi \vee \gamma)$, $(\phi \rightarrow \gamma)$, $(\phi \leftrightarrow \gamma)$:

$\mathcal{M},w \models \alpha \Leftrightarrow \mathcal{M},w \models \psi$ and $\mathcal{M},w \models \gamma$

$\mathcal{M},w \models \alpha \Leftrightarrow \mathcal{M},w \models \psi$ or $\mathcal{M},w \models \gamma$

$\mathcal{M},w \models \alpha \Leftrightarrow$ If $\mathcal{M},w \models \psi$, then $\mathcal{M},w \models \gamma$

$\mathcal{M},w \models \alpha \Leftrightarrow \mathcal{M},w \models \psi$ if and only if $\mathcal{M},w \models \gamma$

For $\alpha = B_S(\psi)$: $\mathcal{M},w \models \alpha \Leftrightarrow \mathcal{M},w' \models \psi$ for all w' such that wRw'

For $\alpha = K_S(\psi)$: $\mathcal{M},w \models \alpha \Leftrightarrow \mathcal{M},w' \models \psi$ for all w' such that wR_Kw'

An $\mathcal{L}_{B,K}$-formula α is said to be *valid in a frame* $\mathcal{F}$ if, for all models $\mathcal{M}$ over $\mathcal{F}$ and all $w \in \mathcal{M}$: $\mathcal{M}, w \models \alpha$. In particular, note that for any frame $\mathcal{F}$, the $\mathcal{L}_{B,K}$-formula $(\psi \rightarrow \gamma)$ is valid in $\mathcal{F}$ just in case for any model $\mathcal{M}$ over $\mathcal{F}$, and any w in $\mathcal{M}$'s domain, $\mathcal{M}, w \models \psi$ only if $\mathcal{M}, w \models \gamma$.[1]

Valid Inferences

Which inferences involving B_S and B_K are validated by the semantics depends in part on the formal properties of R_B and R_K in the relevant models. For example, it is a familiar fact that all formulas of the form $K_S(\alpha) \rightarrow \alpha$ are valid in all frames in which R_K is reflexive (call them R_K-*reflexive* frames). And, conversely, if all formulas of the form $K_S(\alpha) \rightarrow \alpha$ are valid in a class of frames, then in all members of that class, R_K is reflexive. But some inferences are validated regardless of any formal properties

[1]**Proof**: Suppose, for the left-to-right direction, that $(\psi \rightarrow \gamma)$ is valid in $\mathcal{F}$ and that, for some arbitrary model $\mathcal{M}$ and world w: $\mathcal{M}$, $w \models \psi$. Since, by assumption, $(\psi \rightarrow \gamma)$ is valid in $\mathcal{F}$, it follows that $\mathcal{M}$, $w \models (\psi \rightarrow \gamma)$. Therefore, since $\mathcal{M}$, $w \models \psi$ and $\mathcal{M}$, $w \models (\psi \rightarrow \gamma)$, it follows that $\mathcal{M}$, $w \models \gamma$.

For the right-to-left direction, suppose that for any model $\mathcal{M}$ over $\mathcal{F}$, and any w in $\mathcal{M}$'s domain, $\mathcal{M}, w \models \psi$ only if $\mathcal{M}, w \models \gamma$, and let $\mathcal{M}^*$ be an arbitrary model over the frame $\mathcal{F}$. In order to show that $(\psi \rightarrow \gamma)$ is valid in $\mathcal{F}$, it is necessary to show that for all worlds w in the domain of $\mathcal{F}$: $\mathcal{M}^*$, $w \models (\psi \rightarrow \gamma)$. Let w^* be an arbitrary world in the domain of $\mathcal{F}$. If $\mathcal{M}^*$, $w^* \not\models (\psi \rightarrow \gamma)$, then it would have to be because $\mathcal{M}^*, w^* \models \psi$ while $\mathcal{M}^*, w^* \not\models \gamma$. But this is inconsistent with the supposition that $(\psi \rightarrow \gamma)$ is valid in $\mathcal{F}$, which is equivalent to the supposition that $\mathcal{M}, w \models (\psi \rightarrow \gamma)$ for all models $\mathcal{M}$ over $\mathcal{F}$ and all worlds w in the domain of $\mathcal{F}$.

of accessibility relations. The so-called *logical omniscience* properties of B_S and K_S are valid in *all* frames:

$$\textbf{(Omniscience)}\ \gamma \text{ is a logical consequence of } \psi \Rightarrow \ \models (B_S(\psi) \rightarrow B_S(\gamma))$$
$$\gamma \text{ is a logical consequence of } \psi \Rightarrow \ \models (K_S(\psi) \rightarrow K_S(\gamma))$$

This is, in part, a consequence of the fact that, in all models, all the worlds in the range of the accessibility relations corresponding to B_S and K_S are logically consistent and complete. I will present a generalization of the semantics we have so far, which invokes worlds lacking logical consistency and/or completeness—worlds that are logically impossible.

Rantala's Non-Normal Frames

Whereas a model over an ordinary frame $\mathcal{F}$ has a domain W consisting of only logically possible worlds, a model over a *non-normal frame* $\mathcal{F}_\perp$ has non-normal worlds in its domain.[2] A non-normal frame $\mathcal{F}_\perp$ is a quadruple (W, N, R_B, R_K) in which W is a set of worlds and N is a set of possible worlds. The set $W - N$ is the set of non-normal worlds[3]—worlds lacking the property of logical consistency and/or completeness. Both R_B and R_K are subsets of $(W \cup N) \times (W \cup N)$. As usual, a model over the non-normal model is a pair ($\mathcal{F}_\perp$, ϕ) consisting of a non-normal frame $\mathcal{F}_\perp$ and an assignment function ϕ. The assignment function in non-normal models maps pairs of worlds and sentences to truth values, where the values of complex sentences at a possible world w are determined recursively from the values of atomic sentences at w. But the assignment function ϕ is not recursively defined for sentences at non-

[2] See Rantala (1982)

[3] Non-normal worlds are the same as *impossible* worlds, but we'll use 'non-normal' in order to be consistent with the label for non-normal frames.

normal worlds. Rather, the pair (w, α) gets assigned a value directly by ϕ for any $\mathcal{L}_{K,B}$-formula α and $w \in W - N$.

Validity in a non-normal frame $\mathcal{F}_\perp$ is defined exactly as it is defined for validity in ordinary frames. That is, ψ is valid in $\mathcal{F}_\perp$ just in case for all models $\mathcal{M}$ over $\mathcal{F}_\perp$ and normal worlds w (worlds in $W - N$) in $\mathcal{M}$: $\mathcal{M}, w \models \psi$. It is not hard to see why we restrict attention to *normal* worlds; since non-normal worlds are impossible, we do not expect them to be useful in modeling logically valid inferences. The utility of non-normal worlds in a frame $\mathcal{F}_\perp$ is the fact that they can be in the range of $\mathcal{F}_\perp$'s accessibility relations R_B and R_K.[4]

Note that the logical omniscience properties are no longer valid in all frames, if by "all frames" we mean to include non-normal frames. Consider a model $\mathcal{M}$ over a non-normal frame such that its domain contains a pair of worlds (w, w') where w is normal and w' is non-normal. Suppose that, even though γ is a logical consequence of ψ, ψ is assigned the value 1 at w' but γ is assigned the value false at w'. Let the pair (w, w') be a member of $\mathcal{M}$'s accessibility relation for belief R_B and let ψ be assigned the value 1 at every other world that is R_B-related to w. Then we have that $\mathcal{M}, w \models B_S(\phi)$ but not $\mathcal{M}, w \models B_S(\gamma)$ even though $\mathcal{M}, w \models \psi \rightarrow \gamma$.

4.2 Corresponding Closure Conditions

In this section we show that any closure property governing belief must also be valid for knowledge according to the unified bimodal semantics above. We will assume,

[4]This is a somewhat subtle point. Non-normal worlds *do* feature in the definition of logical validity, but only indirectly. We do not say, for example, a sentence S is valid in $\mathcal{F}_\perp$ if it is true in all models over $\mathcal{F}_\perp$ and all worlds in the domain of $\mathcal{F}_\perp$—normal and non-normal. However, non-normal worlds do exist in the domains of non-normal frames. These worlds can be in the range of accessibility relations R_K and R_B, which is how they indirectly feature in the definition of validity

conservatively, that the frames appropriate for modeling knowledge are exactly the R_K-reflexive/transitive/symmetric frames. This assumption is conservative because the conclusion of the argument in this section is constrained by the set of formal properties governing the R_Ks—in general, the stronger our assumptions about the formal properties of the R_Ks, the weaker our results will be. Similarly with the accessibility relations for belief. We assume, conservatively, that the frames appropriate for modeling belief are R_B-serial/transitive/symmetric.[5]

Let α be an $\mathcal{L}_{B,K}$-formula. For any model $\mathcal{M}$, normal or non-normal, there will be a set of worlds w in $\mathcal{M}$'s domain such that $\phi(w, \alpha) = 1$. We use the double bracket notation $[\![\alpha]\!]_{\mathcal{M}}$ to denote this set of worlds. Recall that, by definition, an $\mathcal{L}_{B,K}$-formula α is *valid* in all R_B-serial/transitive/symmetric frames if for every model $\mathcal{M}$ over any such frame and any possible world w in the domain of $\mathcal{M}$, $\phi(w, \alpha) = 1$ (where ϕ is the assignment function associated with the model $\mathcal{M}$). This condition is equivalent to the condition that $[\![\alpha]\!]_{\mathcal{M}}$ is identical to the set of possible worlds in $\mathcal{M}$, for any model $\mathcal{M}$ over a R_B-serial/transitive/reflexive frame. We will say that a frame $\mathcal{F}$ is a Σ-*frame* if its doxastic accessibility relation R_B is serial, transitive, and euclidean and its epistemic accessibility relation R_K is reflexive, transitive, and symmetric. Consider the following assertion:

> $(*)$ If, for any model $\mathcal{M}$ over a Σ-frame and for any $\mathcal{L}$-formulas α, ψ, if $[\![\alpha]\!]_{\mathcal{M}} \subseteq [\![\psi]\!]_{\mathcal{M}}$, then the formulas '$B_S(\alpha) \to B_S(\psi)$' and '$K_S(\alpha) \to K_S(\psi)$' are valid in all Σ-frames.

The proof of $(*)$ is straightforward: If every model $\mathcal{M}$ over a Σ-frame is such that $[\![\alpha]\!]_{\mathcal{M}} \subseteq [\![\psi]\!]_{\mathcal{M}}$, then there cannot be any model $\mathcal{M}'$ over a Σ-frame such the subset of

[5]These assumptions about the formal properties of accessibility relations for knowledge and belief operators are defended in Stalnaker (2006).

possible worlds in $[\![B_S\alpha]\!]_{\mathcal{M}'}$ is not a subset of the set of possible worlds in $[\![B_S\psi]\!]_{\mathcal{M}'}$, otherwise there would have to be a possible world w and a world w' (both in the domain of $\mathcal{M}'$) such that wR_Bw' and $\phi(\alpha, w') = 1$ and $\phi(\psi, w') = 0$ (where R_B and ϕ are the accessibility relation for belief and the assignment function in $\mathcal{M}'$ respectively). But that would mean $w' \in [\![\alpha]\!]_{\mathcal{M}'}$ and $w' \notin [\![\psi]\!]_{\mathcal{M}'}$ contra the assumption that $[\![\alpha]\!]_{\mathcal{M}} \subseteq [\![\psi]\!]_{\mathcal{M}}$ for every model $\mathcal{M}$ over a Σ-frame. The proof that'$K_S(\alpha) \to K_S(\psi)$' is also valid in all Σ-frames is the same.

We will say that an $\mathcal{L}_{B,K}$-formula is Σ-*valid* if it is valid in all Σ-frames. We now have enough machinery to state and prove the first of our two main results:

> **(Closure 1)** For $\mathcal{L}$-formulas α, ψ: If '$B_S(\alpha) \to B_S(\psi)$' is Σ-valid, then '$K_S(\alpha) \to K_S(\psi)$' is Σ-valid.

Here is the proof for (Closure 1): Suppose that '$K_S(\alpha) \to K_S(\psi)$' is not Σ-valid. Then, by $(*)$, it follows that there is some model $\mathcal{M}$ over a Σ-frame such that $[\![\alpha]\!]_{\mathcal{M}} \nsubseteq [\![\psi]\!]_{\mathcal{M}}$. That being the case, we can specify a new model $\mathcal{M}'$ over a Σ-frame based on the model $\mathcal{M}$. Let w be a possible world in the domain of $\mathcal{M}'$ and let R'_B be such that $\{w' : wR'_Bw'\} \subseteq [\![\alpha]\!]_{\mathcal{M}'}$ and $\{w' : wR'_Bw'\} - [\![\psi]\!]_{\mathcal{M}'} \neq \emptyset$. If $\mathcal{M} = ((W, R_B, R_K), \phi)$, set $\mathcal{M}' = ((W, R'_B, R_K), \phi)$. A a graphical representation of the construction of $\mathcal{M}'$ from $\mathcal{M}$ is helpful. Here is a view of the relevant portion of the model $\mathcal{M}$ (without $\{w' : wR_Bw'\}$):

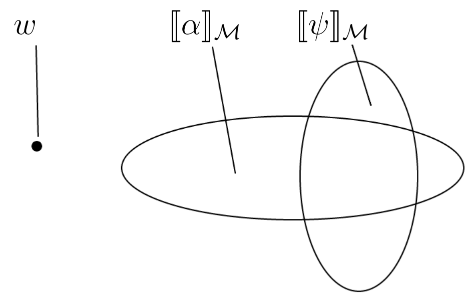

Figure 5

And here is the relevant portion of $\mathcal{M}'$ in which the shaded area represents $\{w' : wR'_Bw'\}$:

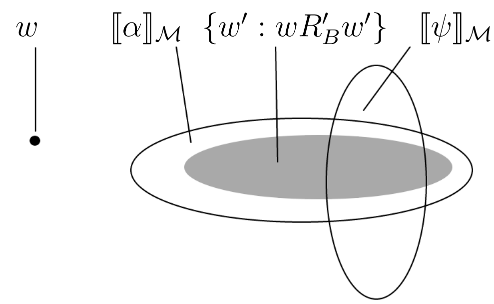

Figure 6

In order to ensure that R'_B is serial, transitive, and euclidean, we stipulate that for any two worlds u and v in the range of R'_B, the pair (u,v) is itself in R'_B. So $\mathcal{M}'$ is a model over a Σ-frame in which $B_S(\alpha)$ is true at the possible world w but $B_S(\psi)$ is not true at w. It follows that the $\mathcal{L}_{B,K}$-formula '$B_S(\alpha) \to B_S(\psi)$' is not Σ-valid, as required.

This result is troubling because there seems to be no pretheoretical motivation for the claim that if a particular inference preserves belief, then it also preserves knowledge. In fact, this has the appearance of a fallacy. If we think of belief as, in some sense, *part* of knowledge, then this would be an instance of the fallacy of composition—the thought that things inherit the properties of their proper parts.[6] But the situation gets worse. With only some superficial qualifications, we can prove the inverse of (Closure 1). Let '@' denote the actual world in the "intended" model $\mathcal{M}_{\mathcal{I}}$ of the language $\mathcal{L}_{B,K}$. We have the following result:

> **(Closure 2)** For $\mathcal{L}$-formulas α, ψ: If '$B_S(\alpha) \rightarrow B_S(\psi)$' is false at @ in $\mathcal{M}_{\mathcal{I}}$ and $@ \in [\![\alpha]\!]_{\mathcal{M}_{\mathcal{I}}}$, then '$K_S(\alpha) \rightarrow K_S(\psi)$' is not Σ-valid.

Here is the proof for (Closure 2): Suppose that '$B_S(\alpha) \rightarrow B_S(\psi)$' is false at @ and $@ \in [\![\alpha]\!]$. By $(*)$ it follows that there is a model $\mathcal{M}$ over a Σ-frame in which $[\![\alpha]\!]_{\mathcal{M}} \nsubseteq [\![\psi]\!]_{\mathcal{M}}$. In fact, this model is $\mathcal{M}_{\mathcal{I}}$ itself. We will construct a model $\mathcal{M}'$ over a Σ-frame based on $\mathcal{M}_{\mathcal{I}}$ where $K_S(\alpha)$ is true at @ but '$K_S(\psi)$' is not. Let R'_K be such that $\{w' : @R'_B w'\} \subseteq [\![\alpha]\!]_{\mathcal{M}'}$ and $\{w' : @R'_B w'\} - [\![\psi]\!]_{\mathcal{M}'} \neq \emptyset$. If $\mathcal{M}_{\mathcal{I}} = ((W, R_B, R_K), \phi)$, set $\mathcal{M}' = ((W, R_B, R'_K), \phi)$. Again, a graphical representation of the construction of $\mathcal{M}'$ from $\mathcal{M}_{\mathcal{I}}$ might be helpful. Here is a view of the relevant portion of the model $\mathcal{M}_{\mathcal{I}}$ (without $\{w' : @R_K w'\}$):

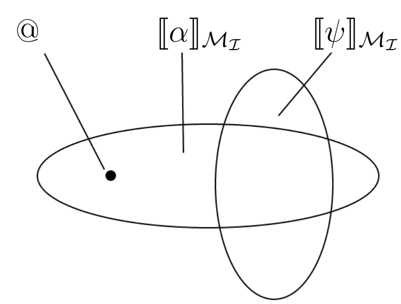

Figure 7

And here is the relevant portion of $\mathcal{M}'$ in which the shaded area represents $\{w' :$

[6]See, for example, Hales (1995) for an argument that the "parts" of knowledge (like belief) do not necessarily inherit the closure properties of knowledge. Brueckner (2004) argues that under relatively modest assumptions, any proposed counterexamples can be explained away.

$@R'_K w'\}$:

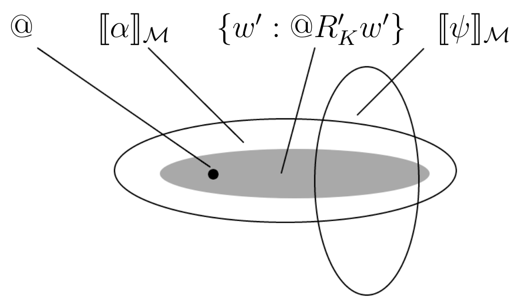

Figure 8

In order to ensure that R'_K is an equivalence relation, we stipulate that for any two worlds u and v in the range of R'_B, the pair (u,v) is itself in R'_B. So $\mathcal{M}'$ is a model over a Σ-frame in which $K_S(\alpha)$ is true at the possible world @ but $K_S(\psi)$ is not true at @. It follows that the $\mathcal{L}_{B,K}$-formula '$K_S(\alpha) \to K_S(\psi)$' is not Σ-valid, as required.

4.3 Generalizations

Instead of discussing particular pairs of $\mathcal{L}$-formulas, we can generalize the discussion to rules of deductive inference that generate new $\mathcal{L}$-formulas from old ones. Let ρ be some rule of inference that operates on individual $\mathcal{L}$-formulas—a single-premise deductive rule. We will say that the belief operator B_S is *closed under* the single-premise rule ρ if the $\mathcal{L}_{B,K}$-formula '$B_S(\alpha) \to B_S(\psi)$' is valid in all Σ-frames whenever the $\mathcal{L}$-formula ψ follows from the $\mathcal{L}$-formula α by a single application of the rule ρ—and similarly for the knowledge operator K_S. The following is an immediate corollary of (Closure1):

(Corollary 1): For any single-premise deductive rule ρ if B_S is closed under ρ, then so is K_S.

To see why (Closure) implies (Corollary 1), assume that K_S is *not* closed under the rule ρ. Then there exists at least one pair of $\mathcal{L}$-formulas α and ψ such that the $\mathcal{L}_{B,K}$-formula '$K_S(\alpha) \rightarrow K_S(\psi)$' is not valid in all Σ-frames. By (Closure 1) it follows that '$B_S(\alpha) \rightarrow B_S(\psi)$' is not valid in all Σ-frames. But that is just to say that B_S is not closed under the rule ρ, as required.

The following generalization of (Closure 2) is similar:

> **(Corollary 2)**: For any single-premise deductive rule ρ, if there exist $\mathcal{L}$-formulas α, ψ such that ψ follows from α by an application of ρ, then if $@ \in [\![B_S(\alpha)]\!]_{\mathcal{M}_\mathcal{I}} - [\![B_S(\psi)]\!]_{\mathcal{M}_\mathcal{I}}$, then K_S is not closed under ρ.

To see why (Corollary 2) follows from (Closure 2), assume that $@ \in [\![B_S(\alpha)]\!]_{\mathcal{M}_\mathcal{I}} - [\![B_S(\psi)]\!]_{\mathcal{M}_\mathcal{I}}$. This is equivalent to the falsity of the formula '$B_S(\alpha) \rightarrow B_S(\psi)$' at @. It follows, by (Closure 2), that '$K_S(\alpha) \rightarrow K_S(\psi)$' is not Σ-valid. But since ψ follows from α by an application of ρ, K_S is not closed under ρ.

As it stands, we still do not have a very general result. After all, these two corollaries only apply to deductive rules that operate on a single $\mathcal{L}$-formula. We can prove an analogous corollary about multi-premise deductive rules. We first need to establish the following generalization of $(*)$.

> **$(*')$** If, for any model $\mathcal{M}$ over a Σ-frame and for any $\mathcal{L}$-formulas $\alpha_1, ..., \alpha_n, \psi$, if $\bigcap_{i \leq n} [\![\alpha_i]\!]_\mathcal{M} \subseteq [\![\psi]\!]_\mathcal{M}$, then the formula '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is valid in all Σ-frames.

We will skip the proofs of $(*')$ because is a routine generalization of the proofs of $(*)$. However, given $(*')$, the following multi-premise version of (Closure 1) follows easily:

(GC 1) For $\mathcal{L}$-formulas $\alpha_1, ..., \alpha_n$, ψ, if the $\mathcal{L}_{B,K}$ formula '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is valid in all Σ-frames then so is the $\mathcal{L}_{B,K}$ formula '$(K_S(\alpha_1) \wedge ... \wedge K_S(\alpha_n)) \rightarrow K_S(\psi)$'.

The proof of (GC 1) is similar to that of (Closure 1): Suppose that '$(K_S(\alpha_1) \wedge ... \wedge (\alpha_n)) \rightarrow K_S(\psi)$' is not Σ-valid. Then, by $(*')$ we have at least one model $\mathcal{M}$ over a Σ-frame in which $\bigcap_{i \leq n} [\![\alpha_i]\!]_{\mathcal{M}} \nsubseteq [\![\psi]\!]_{\mathcal{M}}$. We construct a model $\mathcal{M}'$ based on $\mathcal{M}$. Let w be a possible world in the domain of $\mathcal{M}'$ and let R'_B be such that $\{w' : wR'_B w'\} \subseteq \bigcap_{i \leq n} [\![\alpha_i]\!]_{\mathcal{M}'}$ and $\{w' : wR'_B w'\} - [\![\psi]\!]_{\mathcal{M}'} \neq \emptyset$. If $\mathcal{M} = ((W, R_B, R_K), \phi)$, set $\mathcal{M}' = ((W, R'_B, R_K), \phi)$. A graphical representation of the construction of $\mathcal{M}'$ from $\mathcal{M}$ is helpful. Here is the relevant part of the model $\mathcal{M}$:

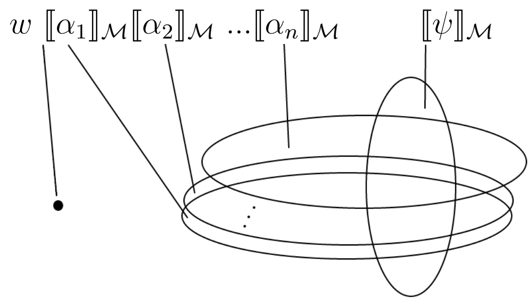

Figure 9

And here is the relevant part of the model $\mathcal{M}'$:

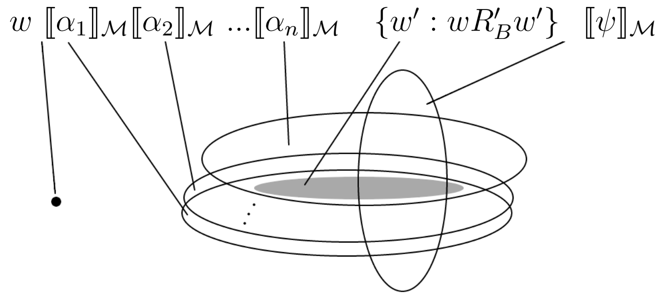

Figure 10

In order to ensure that R'_B is serial, transitive, and euclidean, we stipulate that for any two worlds u and v in the range of R'_B, the pair (u,v) is itself in R'_B. So $\mathcal{M}'$ is a model over a Σ-frame in which $B_S(\alpha)$ is true at the possible world w but $B_S(\psi)$ is not true at w. It follows that the $\mathcal{L}_{B,K}$-formula '$B_S(\alpha) \rightarrow B_S(\psi)$' is not Σ-valid, as required.

Similarly, we have a multi-premise version of (Closure 2). Again we let '@' denote the actual world in the intended model $\mathcal{M}_{\mathcal{I}}$:

> **(GC 2)** For $\mathcal{L}$-formulas $\alpha_1, ..., \alpha_n$, ψ, if the $\mathcal{L}_{B,K}$-formula '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is false at @ in $\mathcal{M}_{\mathcal{I}}$, then the $\mathcal{L}_{B,K}$-formula '$(K_S(\alpha_1) \wedge ... \wedge K_S(\alpha_n)) \rightarrow K_S(\psi)$' is not Σ-valid.

The proof of (GC 2) is a routine generalization of the proof of (Closure 2).

We will say that the belief operator B_S is *closed under* the multi-premise rule ρ if whenever the $\mathcal{L}$-formula ψ follows from the $\mathcal{L}$-formulas α_i by a single application of the rule ρ, the formula '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is valid in all Σ-frames—and similarly for the knowledge operator K_S. We have the following generalization of (GC 1):

> **(Corollary 3)** For any multi-premise deductive rule ρ, if B_S is closed under ρ, then so is K_S.

Suppose that ρ is an n-ary deductive rule and that K_S is not closed under ρ. That means that there are $\mathcal{L}$-formulas $\alpha_1, ..., \alpha_n$, ψ such that '$(K_S(\alpha_1) \wedge ... \wedge K_S(\alpha_n)) \rightarrow K_S(\psi)$' is not valid in all Σ-frames. (GC 1) implies that the sentence '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is not Σ-valid. But that is just to say that B_S is not closed under the rule ρ, as required. Finally, we have a generalization of (GC 2). We will say that the belief operator B_S is *actually closed* under a multi-premise deductive rule ρ if the formula '$(B_S(\alpha_1) \wedge ... \wedge B_S(\alpha_n)) \rightarrow B_S(\psi)$' is true at @ in the intended model $\mathcal{M}_\mathcal{I}$ whenever an $\mathcal{L}$-formula ψ follows from $\mathcal{L}$-formulas α_i by an application of the rule ρ.

> **(Corollary 4)** For any multi-premise deductive rule ρ, if B_S is not actually closed under ρ, then K_S is not closed under ρ.

Suppose that ρ is an n-ary deductive rule and that B_S is not actually closed under ρ. That means that there are $\mathcal{L}$-formulas $\alpha_1, ..., \alpha_n$, ψ such that '$(K_S(\alpha_1) \wedge ... \wedge K_S(\alpha_n)) \rightarrow K_S(\psi)$' is not valid in all Σ-frames (since the intended model $\mathcal{M}_\mathcal{I}$ is a model over a Σ-frame). (GC 2) implies that the sentence '$(K_S(\alpha_1) \wedge ... \wedge K_S(\alpha_n)) \rightarrow K_S(\psi)$' is not Σ-valid. But that is just to say that K_S is not closed under the rule ρ, as required.

Neither (Corollary 3) nor (Corollary 4) are very attractive results. They establish that on a generalization of the accepted model-theoretic semantics for belief and knowledge, we cannot validate closure properties of belief without thereby being committed to the same closure properties for knowledge, and vice versa. Since there seems to be no pre-theoretical motivation for this result, it would be convenient if there were an easy fix. In the next section I show that in principle an ad hoc fix is available which allows for denying part of the proof of (Corollary 3). However, this requires some substantive revision of the semantics and is a non-starter for undercutting the proof of (Corollary 4).

4.4 Restricting Metalinguistic Quantification

We have so far made use of the assumption that nothing restricts the domain of the quantification on the right side of the truth-at-a-model/world conditions of knowledge-ascribing formulas of the form $K_S(\psi)$:

$$(K_S) \qquad \mathcal{M},w \models K_S(\psi) \;\Leftrightarrow\; \mathcal{M},w' \models \psi \text{ for all } w' \text{ such that } wRw'$$

That is, nothing restricts the domain of quantification except for the restriction imposed by the model's doxastic accessibility relation R_B. Of course, there are restrictions on the relation of R_K (i.e. in any Σ-frame, R_B must be reflexive, transitive, and symmetric), and these in turn impose restrictions on the range of R_K from any given possible world w. Any such restriction on the range of R_K in turn imposes a restriction on the domain of quantification on the right-hand side of K_S. In the proof of (Closure 2), back in sect. 2, we were careful not to violate any formal constraints on the relation R_K in any Σ-frame. But this raises the question, what constraints *could be* imposed on an epistemic accessibility relation R_K, such that the proof of (Closure 2) is undercut?

There is another way of posing the same question that may have slightly more intuitive appeal. Let S's *epistemic space at* w be the set of all worlds that are R_K-related to a possible world w (remember that R_K is always defined relative to an agent

S). We can think of S's epistemic space at w as the set of worlds that are, in some sense, *compatible* with what S believes at w.[7] So, of course, among the worlds that are excluded from S's doxastic space at w are all those worlds that are incompatible with, or ruled out by, what S believes at w. But the agent S can have different beliefs across different possible worlds and across different models. This raises the following question: what sorts of worlds are consistently ruled out as being members of S's doxastic space in *any* model $\mathcal{M}$ and *any* possible world w? That is, what sort of world is, independent of an S's beliefs at a particular world, a candidate for membership in S's doxastic space. So far we have not imposed any special restrictions in this regard. Suppose, then, that we impose the following restriction on models in which 'ρ' denotes some deductive rule of inference:

> **($\boldsymbol{K_S}$-restriction)** For any model $\mathcal{M}$ and possible world w in $\mathcal{M}$'s domain, if $w' \in \{w' : wR_K w'\}$, then w' is ρ-complete.

Call any Σ-frame in which (K_S-restriction) holds a Θ-*frame*. If indeed (K_S-restriction) is a proper restriction of S's epistemic space—if there is no way for a world to be epistemically accessible for S without being ρ-complete—then the class of frames relevant for modeling knowledge is just the class, or some further restricted subclass of, Θ-frames. Here is the point: (Closure 3) does not admit of a proof of the sort given for (Closure 2) in sect. 2.

> **(Closure 3)** For $\mathcal{L}$-formulas α, ψ: If '$K_S(\alpha) \rightarrow K_S(\psi)$' is Θ-valid, then '$B_S(\alpha) \rightarrow B_S(\psi)$' is Θ-valid.

In fact, (Closure 3) is false, and it is not hard to see why. We have imposed an ad hoc constraint on doxastic accessibility that need not apply to *epistemic* (knowledge-

[7]See Hintikka (1969) for details on the compatibility relation.

based) accessibility. In particular, if ψ follows from α by an application of the rule ρ, then the antecedent of (Closure 3) is true, while its consequent is false if there are any Θ-models whose domains include ρ-incomplete worlds.

Before pursuing this option, we should ask why (Closure 1)/(Closure 2) and (Corollary 3)/(Corollary 4) are such unattractive results. It was shown that under very weak assumptions, the worlds-based models of propositional attitudes are incapable of making fine distinctions between various attitude predicates. But are these genuine distinctions? Recall from the opening section the "candidate" closure principles for belief and knowledge:

> **($\mathbf{MP}_B$)** For any rational agent S, if S believes that A, and S believes that A entails B, then S believes that B.
>
> **($\mathbf{MP}_K$)** For any rational agent S, if S knows that A, and S knows that A entails B, then S knows B.

We skipped giving an argument for the truth of (MP_B) and the falsity of (MP_K) for the sake of avoiding some unnecessary philosophical scruples prior to more important material. But, it is precisely the fact that there are cases such as (MP_B) and (MP_K) illustrating the asymmetry of closure conditions between distinct attitude predicates that make (Closure 1)/(Closure 2) and (Corollary 3)/(Corollary 4) troubling in the first place; so an intuitive argument for this asymmetry of closure conditions would be useful.

The principle (MP_B) is about *ideally rational* epistemic agents. Since its scope is limited to such agents, the counterexamples typically proposed would not be relevant (i.e. imagine an agent who believes a conditional together with its antecedent, but fails to believe its consequent). This is not to say that it is impossible for Vann

Mcgee-type rational agents to believe both P and "if P, then Q" while rejecting Q.[8] It is possible, after all, that Vann Mcgee is mistaken about the logic of "if, then" while maintaining rationality. But modeling the propositional attitudes of ideally rational agents is always relative to *some* presupposed background logic. If, when all truths are revealed, Vann Mcgee turns out to be correct about the logic of "if, then" so that modus ponens turns out to be an invalid inference form, this would entail that the presupposed background logic in our models should change to accommodate this fact. As of now, since Vann Mcgee is more or less a voice in the wilderness concerning the invalidity of modus ponens, a model of epistemic rationality should presuppose a background logic according to which rational belief patterns according to modus ponens. In particular, assuming that modus ponens is a valid inference form, Vann McGee's doxastic behavior cannot be modeled as ideally rational if he indeed believes a conditional and its antecedent, but rejects its consequent. In general, this explains how (MP_B) could be true without denying the rationality of deviant logicians nor denying that they believe what they believe—deviant logicians can be considered rational "from the outside" but not modeled as rational in a formal framework which presupposes a fixed background logic.

Given that we are limiting attention to rationally ideal epistemic agents, the argument for the falsity of (MP_K) is not as simple as pointing to cases in which an agent knows a conditional and its antecedent, but fails to know its consequent in virtue of failing to believe it. Since, we will assume, knowledge entails belief, knowing a conditional and its antecedent entails believing them, but then by (MP_B), knowing a conditional and its antecedent entails believing the consequent. If (MP_K) if false, it

[8]See McGee (1985).

is not because belief is not closed under believed entailment.

Consider a sequence of, say, 100 sentences $\phi_1, ..., \phi_{10}$. The ϕ_is jointly entail the conjunction $\wedge\phi_i$. It is possible that some agent S knows all of the ϕ_is individually without knowing their conjunction. Unless one is willing to hold that knowledge entails absolute certainty, the justification for S's belief in the ϕ_is can be less than perfect. An adaptation of the preface paradox illustrates this point. In Makinson (1965), the preface paradox is presented as a puzzle about rational belief. For any particular sentence in a book, the author, having been sufficiently careful and reflective in writing, justifiably believes that that sentence is true. Nevertheless, the author is also rational enough to hedge up front by conceding that not every sentence in the book is true. That is, the author justifiably believes that not every sentence in the book is true. This shows that justifiable belief does not seem to be closed under the rule of conjunction. Suppose, moreover, that every sentence in the text is true so we can say for any particular sentence in the book, the author *knows* of that sentence that it is true. Since conjunction rule does not preserve justifiable belief, the author does not know that every sentence in the book is true—even though, by assumption, every sentence in the book is true and known by the author to be true. The following principle is false:

(1) **(Multi-Premise Closure)** If, while knowing $p_1, ..., p_n$, S believes q on the basis of deducing it from the p_is, the S knows q.

In other words, knowledge is not preserved under multi-premise entailment.

This is the sort of consideration suggesting that not all propositional attitude predicates share the same set of closure properties. One might object by pointing out that knowledge is not preserved by the rule of conjunction only insofar as rational belief is not preserved by rule of conjunction. So, the preface paradox cannot be used to show how the closure properties of knowledge and idealized belief predicates can diverge. But this is a confusion. Our assumption is that *idealized* belief is belief that patterns according to logical consequence in the following sense: ideally, an agent does not believe the premises of a valid argument without believing its conclusion.[9] Note that this accommodates non-monotonicity—if an ideal agent comes to disbelieve the conclusion of a valid argument, then that agent will come to disbelieve at least one premise of the argument. This is using the notion of *ideal belief* not in the sense of having perfect justification. Ideal belief in the relevant sense is just the sort of belief that patterns according to logical consequence. If the relevant sense of ideal belief meant something like having perfect justification, then why not also require that ideal belief is also *true* belief? That is, the sense in which ideal belief entails certainty is also the sense in which ideal belief simply entails truth. Of course, we want to distinguish ideal belief as such from perfectly rational and true belief. So it is consistent to say that ideal belief is preserved, for example, under the rule of conjunction, while knowledge is not. In particular, if the author in our adaptation of the preface case is taken to be an ideal believer, then the author's belief in each individual sentence would entail a belief in the conjunction of all the sentences—which is not to say that any of these beliefs are perfectly justified (otherwise they would have to be true). And this is consistent with the claim that knowledge fails to be closed under the rule of conjunction.

Moreover, the asymmetry of closure properties between knowledge and belief is not relegated to multi-premise arguments. We can re-use the previous argument, which

[9] Field (2009) argues that this is part of the normative role of logic in constraining rational belief.

depends on preface-type cases, in order to show that (MP_K)—which say knowledge is closed under single-premise known entailment—is false. In particular, since we are supposing that we are dealing with ideally rational agents who believe the logical consequences of their beliefs, the following more precise version of (MP_K) is what is at issue:

(2) **(Single-Premise Closure)** If, while knowing p, S believes q by inferring it from the knowledge that p entails q, then S knows q.

This principle turns out to be equivalent to the assumption that knowledge is closed under multi-premise entailment if we assume that the relevant agent knows some basic facts about entailment and conjunction. But if that's the case, then the reasons for thinking that knowledge is not closed under multi-premise entailments are also reasons for thinking that knowledge is not closed under *single*-premise known entailment—which suggests that (MP_K) is false.

(Multi-Premise Closure) entails (Single-Premise Closure) because the latter is a special case of the former. We will show that the converse is also true under weak assumptions. Here is the reasoning: Suppose that S knows the following logical truth.

(3) $(p_1 \to (p_2 \to (p_3 \to ...(p_n \to (p_1 \wedge ... \wedge p_n)...)$

Suppose also that S knows each of the p_i's and that, from (3) together with the knowledge that p_1, S infers the consequent of (3), namely:

(4) $(p_2 \to (p_3 \to (p_4 \to ...(p_n \to (p_1 \wedge ... \wedge p_n)...)$

By (Single-Premise Closure), it follows that S knows (4). Now repeat this reasoning n times. It follows that S knows the conjunction $(p_1 \wedge ... \wedge p_n)$.[10]

[10]This argument was developed over a conversation with Jonathan Livengood.

The only assumption made—in addition to the assumption of (Single-Premise Closure) for the sake of reductio—is that the agent S knows (3) and believes the result of repeated applications of modus ponens. At each step (Single-Premise Closure) guarantees that S knows the conclusion of the modus ponens inference. If cases like the preface paradox show that (Multi-Premise Closure) is false, the argument here shows that these cases also falsify (Single-Premise Closure). Although the argument does not specify *which* of the n applications of (Single-Premise Closure) yields a falsity from a truth, we know something has gone wrong—(Single-Premise Closure) together with the harmless auxiliary assumption that S knows (3) generates the implausible instances of (Multi-Premise Closure).

If the arguments in this section are sound, the fact that familiar worlds-based semantics for propositional attitudes makes all attitude predicates the same with respect to their closure properties is not a happy result. Is it possible to solve this problem by stipulating from the outset that there will be a restriction on the sets of worlds that are in the range of epistemic (knowledge-based) accessibility relations? The beginning of this section describes the way this solution would work. It would be constitutive of epistemic accessibility relations that, in any frame $\mathcal{F}$ and model $\mathcal{M}$ over $\mathcal{F}$ there are certain worlds in the domain of $\mathcal{F}$ that are barred from membership in the range of the epistemic accessibility relation of $\mathcal{F}$. This would ensure "from the outside" that there are certain inference forms that preserve knowledge that need not preserve belief. For example, we could stipulate that there can be no model $\mathcal{M}$ over frame $\mathcal{F}$ such that there are members of the range of $\mathcal{F}$'s accessibility relation which, in $\mathcal{M}$, support the truth of ϕ and $(\phi \rightarrow \psi)$ without supporting the truth of ψ. Under such a stipulation, we guarantee that knowledge predicates obey (MP_K)

without thereby guaranteeing that belief predicates obey (MP_B).

Of course, this is precisely the sort of asymmetry between belief and knowledge that we would want. The arguments developed above were aimed at establishing the existence of some inference patterns that are belief-preserving without also being knowledge preserving. A formal stipulation that guarantees there are no inference patterns preserving knowledge that aren't also belief preserving would help. That is, the falsity of (Closure 3) is precisely what we would want in order to solve the present problems.

The worry is that this "solution" is more revisionary than it might seem. Ordinarily the closure properties of modal operators fall out of set-theoretic properties of accessibility relations which are supposed to hold across a class of frames. For example, the sentence $K_s(\phi) \rightarrow \phi$ is valid in all frames in which the doxastic accessibility relation is reflexive. The expression K_s is being interpreted as a knowledge operator in all such frames. In order to study and generate interesting results about the logic of knowledge, we restrict attention to frames whose epistemic accessibility relations are reflexive. Contrast that with the present strategy for forcing a divergence between closure properties of various propositional attitude predicates. Not only are we supposed to restrict our attention to a particular subclass of frames—i.e. restricting attention those those frames whose epistemic accessibility relations are equivalence relations, and whose doxastic accessibility relations are serial, transitive, and euclidean—we are also suppose to restrict attention to particular models defined *over* such frames. Subclasses of models defined over particular types of frames become the primary object of study. This sort of solution is revisionary insofar as it is paradigm-changing. But insofar as it is paradigm-changing it is also ad hoc. Interesting logical properties of modal operators track differences at the level of frames, not models. The only possible difference between any two models defined over a particular frame is that they assign different truth values to sentence-world pairs. Restricting attention to particular models within a fixed class of frames is analogous to restricting attention to some but not all rows of a truth table—this is analogous to claiming that material conditionals will always express tautologies whenever we are restricting attention to rows in which the antecedent is true and the consequent is false.

Interestingly, though, there is *no* such ad hoc strategy that will work for falsifying the converse of (Closure 3):

> **(Closure 4)** For $\mathcal{L}$-formulas α, ψ: If '$B_S(\alpha) \rightarrow B_S(\psi)$' is Θ-valid, then '$K_S(\alpha) \rightarrow K_S(\psi)$' is Θ-valid.

We have not seen any proposed counterexamples to (Closure 4), but in principle there is no obvious reason that it should hold.[11] Here is why the present strategy of imposing outside restrictions on the range of accessibility relations will not work to falsify (Closure 4). Suppose that we impose the following restriction on models in which 'ρ' denotes some deductive rule of inference:

> **($\boldsymbol{B_S}$-restriction)** For any model $\mathcal{M}$ and possible world w in $\mathcal{M}$'s domain, if $w' \in \{w' : wR_Bw'\}$, then w' is ρ-complete.

Now, since knowledge entails belief, in every model, all epistemically accessible worlds will be doxastically accessible—every world that is compatible with one's knowledge will thereby be compatible with one's beliefs (not vice versa unless all of ones' beliefs count as knowledge). So, the following subset relation obtains:

[11] cf. footnote 6.

(5) $\{w' : wR_Bw'\} \subseteq \{w' : wR_Kw'\}$

The sentence $K_S(\phi) \to B_S(\phi)$ will be valid in all frames that obey (5), and any model over a frame that obeys (5) will be a model that makes the sentence $K_S(\phi) \to B_S(\phi)$ true. In order to falsify (Closure 4), we would want for there to be some model $\mathcal{M}$ over a frame $\mathcal{F}$ (which obeys (5) such that the doxastically accessible worlds that are ρ-incomplete while all epistemically possible worlds are ρ-complete. But, since that would contradict (5), the strategy of stipulating ad hoc restrictions on the range of accessibility relations could only work if there were frames in which (5) is false, in which case the sentence $K_S(\phi) \to B_S(\phi)$ would be invalid.

4.5 Possible Reactions

The results of the previous section are unattractive for the prospects of modeling distinct propositional attitudes with worlds-based models. In previous chapters, the aim was to model propositions themselves using appropriate partitions of the set of possible worlds—a project that is orthogonal to the project of modeling propositional attitudes. If the goal is to use formal machinery to predict the beliefs, inferences, and belief updating behavior of idealized rational agents, then a definition of propositional content which meets intuitively correct individuation constraints is not sufficient. For example, the definition of propositional content in the previous chapter makes relatively few predictions about the doxastic properties of idealized *rational* agents—it does not even predict that epistemic agents must believe the basic logical truths. This is not a shortcoming of the definition; it reflects the fact that a predictive model of rational belief is different from a theory of propositional content.

Perhaps it is a mistake to think that logical laws should play such a central role in modeling rational epistemic agents. If, contra the proposal in the previous chapter, propositions are maximally fine-grained so that no two sentences express the same one, then a relational account of the propositional attitudes would not preclude rational agents from holding radically inconsistent combinations of beliefs. So, if a fixed set of logical laws should not be taken to govern rational belief, the results of the previous section would not undermine the prospects of modeling rational epistemic agents using modal semantics after all.

One reason to think that rational belief is not constrained by logical laws comes from Williamson (2008). The thought that logical laws govern rationality depends on the intuition that, for some sentences (or inferences between sentences), understanding entails assent—not just for normatively idealized agents, but for *any* linguistically competent agent. Williamson's strategy for resisting this claim is to consider individual examples of such sentences (and inferences), and to show how it is possible to have understanding without assent. Here is one example to get a feel for his strategy: Consider the sentence 'All vixens are vixens'. At first blush, it seems that there is no way to understand it without assenting. This sentiment is expressed in the following principle:

(6) Necessarily, whoever understands the sentence"Every vixen is a vixen" assents to it.

Williamson asks us to consider a person, Peter, who withholds assent to 'Every vixen is a vixen' because Peter thinks that the determiner 'every' carries existential import. That is, he thinks that the sentence 'Every vixen is a vixen' entails that there is at least one vixen. Peter also buys into a conspiracy theory according to which there are no foxes. Because of this, he does not assent to 'There is at least one vixen'—nor does he assent to anything he thinks implies it. In particular, Peter does not assent to 'Every vixen is a vixen.' In a somewhat similar case, Stephen withholds assent to 'Every vixen is a vixen', but because of matters having to do with the vagueness of the predicate 'vixen.' Stephen thinks that there are intermediate species between foxes and non-foxes such that predicating vixen to the females of such species yields neither a truth nor a falsity. Steven also accepts strong three-valued logic K_3 in Kleene (1938), so he does not assent to 'Every vixen is a vixen.'

The question is whether Peter and Stephen suffer from linguistic incompetence, or whether they simply hold false views about a sentence they both genuinely understand. The fact that (1) entails that they do not understand the sentence 'Every vixen is a vixen' is, according to Williamson, a counterintuitive result. After all, they are both clearly competent in the use of the constituent expressions 'Every', 'vixen', and 'is a'. A better explanation, according to Williamson, is that Peter and Steven's rogue metalinguistic commitments about the logic of 'every' is a mistake that sometimes propagates through to their non-metalinguistic judgments. These rogue metalinguistic judgments are what explain their unwillingness to assent to 'Every vixen is a vixen', not a failure to understand the sentence. In Williamson's words:

> Their non-metalinguistic unorthodoxy as to when every F is a G is not ultimately derived by semantic descent from metalinguistic orthodoxy as to when "Every F is a G" is true; rather their metalinguistic unorthodoxy is ultimately derived by semantic ascent from their non-metalinguistic unorthodoxy. (Williamson (2008, 90))

This scenario is intended to show that a linguistically competent agent can hold a false belief (perhaps irrationally) about a paradigm analytic truth. Williamson goes on to provide a similar analysis of cases of assent to the premises of paradigm classically-valid inferences, modus ponens and conjunction elimination, without assent to their conclusions. That is, a similar argument is given against the truth of the following principles:

(7) Necessarily, whoever understands and assents to a sentence of the form 'If P, then Q' together with P assents to Q.

(8) Necessarily, whoever understands and assents a sentence of the of the form 'P and Q' assents to both P and Q individually.

How does this help to loosen up the connection between logic and rational belief? There are three steps. First, if Williamson is correct, then it is possible to genuinely believe logical absurdities, which includes believing the premises of simple classically valid arguments without believing the corresponding conclusions. The second step is to notice that the cases Williamson uses to illustrate this point are cases in which there is nothing particularly *irrational* about the relevant beliefs. In the first case, Peter rejects the sentence 'Every vixen is a vixen' because of a rather sophisticated belief about the logic of 'every'. For one to hold, with Aristotle, the view that 'every' carries existential commitment certainly need not entail that one is irrational—even if we suppose that this semantic account of 'every' is, in fact, incorrect. In general, false beliefs about the semantic properties of common expressions does not entail irrationality. One might point out that Peter's acceptance of the conspiracy theory according to which there are no female foxes seems irrational. Fair enough. But this kind of irrationality—concerning matters of fact, rather than matters of logic—is not what's relevant here. Steven's belief that borderline cases lack truth values and that K_3 gets the right results with respect to truth value gaps is similar. That is, one's commitment to this view of vague predicates does not in itself seem to entail that one is irrational. The third step is to recognize that not only are these *not* cases of irrationality, but they are illustrations of a kind of ideal rationality. One's view of rational belief should be consistent with the fact that rational agents can have rational disagreements over logical matters.

Does it follow that the results of the previous sections are unimportant since models of rational belief imposing logical laws are mistaken? Not exactly. For one reason, Williamson's arguments seem to presuppose an inferentialist account of linguistic understanding, according to which competence in the *use* of some expression is sufficient for *understanding* that expression. But, of course, it is not difficult to imagine cases in which a person competently uses an expression without understanding it. Moreover, the cases used to illustrate the possibility of genuine belief in logical absurdities seem to be cases in which rogue metalinguistic commitments are precisely what contaminate an agents' otherwise competent use of an expression. Precisely the cases used to illustrate the possibility of logically absurd beliefs are those cases in which the agent displays some lack of competence with the relevant expressions.

The other option is to accept that coarse-grained possible worlds propositions are hyper-idealized analogues of genuine propositions of the sort that ordinary agents stand in cognitive relations with. They are hyper-idealized not just because of omniscience phenomena, but also because they determine that *ideally* there is no difference between the closure properties of knowledge and belief predicates (or any other propositional attitude predicate with the same formal treatment as knowledge and belief predicates). This sort of parity between distinct propositional attitude predicates is an unattractive artifact of the formal analysis, but this leaves open the question of whether possible worlds propositions are at least idealized approximations of genuine content. Why, and under what conditions, would it follow from the fact that an idealized model of some phenomenon only approximates the empirical phenomenon, that the objects involved in the model cannot be an idealized analogue of the objects under investigation? After all, isn't idealization of some sort or other, to some degree or other, just a part and parcel theoretical modeling?

From the mere fact that laws of Newtonian mechanics are conditional on idealizations such as frictionless planes and rigid bodies, it does not follow that the real physical world with its imperfectly smooth planes and non-rigid bodies is not the real subject under investigation—rather, idealized physical systems are idealized analogues of real physical systems. In the same way, intensions are to be seen as idealized analogues of propositions. The degree to which a semantic theory only applies to normatively ideal agents is precisely the degree to which its intensions are merely approximations of real propositional contents. While the real propositional contents are the things we actually believe, know, doubt, assert, etc., there is no reason to think that an intensional semantic theory of these attitudes is not dealing with models of real propositional contents.

Since this is not the place to address the question of what justifies different sorts of idealization in science, I will simply focus on two of the most widely discussed modes of idealization. I follow Weisberg (2007) in the use of labels *Galilean* and *minimalist* idealization. This response is no doubt too quick to do justice to the complexities of idealization in science, but it is enough to get a general impression of why I think this cannot not a good route for defending maximally ideal theories of propositional content.

Galilean idealization is done with the intent of simplifying theories for practical computational purposes. Since modeling real-world phenomena can easily become computationally intractable, it is useful and justifiable to stipulate some falsities for the sake of gaining computational tractability. This pragmatic justification is supplemented by the fact that as our means of computation increases with technology, we can afford corresponding de-idealizations of the theory. With respect to theories of belief and propositional content, the suggestion is that perhaps idealizations can be given a broadly Galilean justification. But clearly a Galilean justification for idealization does not apply in this case. A defender of idealized coarse-grained propositions is not idealizing away irrational epistemic agents for the sake of gaining anything like computational tractability. The defender of coarse-grained propositions has to idealize away from irrational agents in order to be able to make any non-trivial predictions about belief—computationally tractable or not.

Another mode of idealization that might be more relevant is minimalist idealization—according to which false assumptions are made about irrelevant factors in order to isolate the particular phenomenon that is up for explanation. Consider, for example, the Ising model, according to which atoms and molecules are crudely represented as unanalyzed points which can be in one of two states. If nothing more about the fine structure of atoms and molecules is needed for an explanation a particular phenomenon of interest, then it is justifiable to abstract away from such detail.[12].

Is the minimalist mode of idealization a suitable picture of what is happening in idealized theories of belief and propositional content? The suggestion would be that the fact that there are irrational epistemic agents because of irrelevant matters

[12]The version of minimalism in Strevens (2004) amounts to a model according to which it is justifiable to ignore anything that does not count as a "difference-making" factors

having to do with psychological limitations or whatever, means that it is justifiable to (falsely) assume, for the sake of isolating belief and propositional content as such, that such epistemic agents do not exist. So, just as idealized mechanical theories use frictionless planes, idealized belief theories invoke epistemic agents who lack "cognitive friction."

The trouble is that this appeal to minimalist idealization justifies far more than intended. In fact, if we are abstracting away from all matters irrelevant to belief and content as such, then why not ignore all contingent limitations for the sake of isolating the target phenomenon? It is not as if there is a relevant difference between, on the one hand, the psychological factors leading to logically inconsistent or incomplete belief sets, and, on the other hand, the fact that there is only a finite amount of time before the impending heat death of the universe (and, not to mention, a finite upper bound on human brain's computational capacity). So, what justifies abstracting away from the former but not from the latter? If anything, contingent psychological facts about humans are more relevant to belief as such than are these nebulous facts that make us less than universal Turing machines.

But of course, idealizing away any of these constraints on belief, the constraints determining that we are less than perfect epistemic agents, results in a theory according to which epistemic agents are logically omniscient. The defenders of such theories are not my target in this chapter.[13]

[13]cf. chapter 1 above.

Chapter 5

Bibliography

Barwise, J. and J. Perry (1983). Situations and attitudes.

Bjerring, J. C. (2013). Impossible worlds and logical omniscience: an impossibility result. *Synthese 190*(13), 2505–2524.

Bjerring, J. C. and W. Schwarz (2014). Granularity problems.

Brogaard, B. and J. Salerno (2013). Remarks on counterpossibles. *Synthese 190*(4), 639–660.

Brueckner, A. (2004). Strategies for refuting closure for knowledge. *Analysis 64*(4), 333–335.

Cresswell, M. J. (1985). Structured meanings.

Davidson, D. (1982). Rational animals. *dialectica 36*(4), 317–327.

Dretske, F. (1981). Knowledge and the flow of information.

Dretske, F. (2008). Epistemology and information. *Philosophy of information. Elsevier, Amsterdam*, 29–48.

Elbourne, P. (2010). Why propositions might be sets of truth-supporting circumstances. *Journal of philosophical logic 39*(1), 101–111.

Field, H. (2009). What is the normative role of logic? In *Aristotelian Society Supplementary Volume*, Volume 83, pp. 251–268. Wiley Online Library.

Fodor, J. A. (1990). Psychosemantics, or, where do truth conditions come from?

Frege, G. (1906). A brief survey of my logical doctrines. *H. Hermes, F. Kambartel, and Friedrich Kaulbach (eds and trans.). Frege: Posthumous Writings. Oxford: Basil Blackwell 1979*, 197–202.

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 289–311.

Grice, H. (1975). Logic and conversation'in p. cole and j. morgan (eds.) syntax and semantics volume 3: Speech acts.

Hales, S. D. (1995). Epistemic closure principles. *The Southern Journal of Philosophy 33*(2), 185–202.

Halpern, J. Y. and R. Pucella (2007). Dealing with logical omniscience. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pp. 169–176. ACM.

Heim, I. and A. Kratzer (1998). *Semantics in generative grammar*, Volume 13. Blackwell Oxford.

Hintikka, J. (1962). *Knowledge and Belief.*

Hintikka, J. (1969). Semantics for propositional attitudes. In *Models for modalities*, pp. 87–111. Springer.

Hintikka, J. (1979). Impossible possible worlds vindicated. In *Game-Theoretical Semantics*, pp. 367–379. Springer.

Jago, M. (2012). Constructing worlds. *Synthese 189*(1), 59–74.

Jago, M. (2013). Impossible worlds. *Noûs*.

Kaplan, D. (1999). The meaning of ouch and oops: Explorations in the theory of meaning as use. *Manuscript, UCLA*.

King, J. C. (2007). The nature and structure of content.

Kleene, S. C. (1938). On notation for ordinal numbers. *The Journal of Symbolic Logic 3*(4), 150–155.

Krakauer, B. (2013). What are impossible worlds? *Philosophical studies 165*(3), 989–1007.

Kriegel, U. (2013). The phenomenal intentionality research program. *Phenomenal intentionality*, 1–26.

Kripke, S. A. (1972). *Naming and necessity.* Springer.

Kripke, S. A. (1979). *A puzzle about belief.* Springer.

Lewis, D. (1972). *General semantics.* Springer.

MacFarlane, J. G. (2000). *What does it mean to say that logic is formal?* Ph. D. thesis, Citeseer.

Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 205–207.

McCarthy, T. (1981). The idea of a logical constant. *The Journal of Philosophy*, 499–523.

McGee, V. (1985). A counterexample to modus ponens. *The Journal of Philosophy*, 462–471.

Mendelovici, A. (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical studies 165*(2), 421–443.

Moltmann, F. (2003). Propositional attitudes without propositions. *Synthese 135*(1), 77–118.

Murray, D., J. Sytsma, and J. Livengood (2013). God knows (but does god believe?). *Philosophical studies 166*(1), 83–107.

Nolan, D. et al. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic 38*(4), 535–572.

Peacocke, C. (1976). What is a logical constant? *The Journal of Philosophy*, 221–240.

Priest, G. (1987). In contradiction. a study of the transconsistent, vol. 39 of nijhoff international philosophy series. *Dordrecht, Martinus Nijhoff, The Hague*.

Priest, G. (2005). Towards non-being: The logic and metaphysics of intentionality.

Quine, W. V. (1956). Quantifiers and propositional attitudes. *the Journal of Philosophy*, 177–187.

Quine, W. V. O. (1969). *Ontological relativity and other essays.* Number 1. Columbia University Press.

Quine, W. V. O. (1986). *Philosophy of logic*. Harvard University Press.

Rantala, V. (1982). Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica 35*, 106–115.

Richard, M. (1983). Direct reference and ascriptions of belief. *Journal of Philosophical Logic 12*(4), 425–452.

Russell, B. (1903). *The principles of mathematics*. WW Norton & Company.

Russell, B. (1912). (1912) the problems of philosophy. *London: Home University Library*.

Russell, B. (1913). *Theory of knowledge: The 1913 manuscript*, Volume 7. Psychology Press.

Russell, B. (1918). The philosophy of logical atomism. *The Monist*, 345–380.

Salmon, N. U. (1986). Frege's puzzle.

Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics 15*, 47–87.

Stalnaker, R. (1984). Inquiry.

Stalnaker, R. (1986). Possible worlds and situations. *Journal of Philosophical Logic 15*(1), 109–123.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical studies 128*(1), 169–199.

Stalnaker, R. C. (1976). Possible worlds. *Noûs*, 65–75.

Strevens, M. (2004). The causal and unification approaches to explanation unified?causally. *Noûs 38*(1), 154–176.

Von Fintel, K. and I. Heim (2002). Lecture notes on intensional semantics. *ms., Massachusetts Institute of Technology*.

Waismann, F. (1979). Wittgenstein and the vienna circle: conversations.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 639–659.

Williamson, T. (2008). *The philosophy of philosophy.* John Wiley & Sons.