THE UNIVERSITY OF CHICAGO

MEANING AND THE WORLD

A DISSERTATION SUBMITTED TO
THE FACULTY IN THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY
RYAN SIMONELLI

CHICAGO, IL
AUGUST 2022

Contents

List of Figures

Acknowledgments

Many conversations with many people have shaped the way I think about things discussed here. Without those conversations and, more importantly, those people, this dissertation would not be what it is. So, insofar as anything good is done in this dissertation at all, many thanks are in order.

First off, I'd like to thank my committee members: Michael Kremer, Malte Willer, Jim Conant, Bob Brandom, and Jason Bridges. Michael and Malte were everything one could ask for in a pair of dissertation advisers: available, encouraging, supportive, and, perhaps most importantly for an advisee like me, clear-headed and pragmatic. Michael's Sellars course was what first really exposed me to many of the Sellarsian lines of thought pursued here, and taking courses with Malte, realizing how rich the world of contemporary semantic theorizing is and can potentially be, is what led me to pursue these lines of thought in the way I do here. Though Jason played a more back-seat role in advising, I've benefited greatly from each and every conversation I've had with him. Among my committee members, the two who've exerted the most influence on this project (and on my thinking in general) are, without a doubt, Bob and Jim. The influence of Bob's work is, of course, quite explicit in the text, but Bob's personal guidance and enormous intellectual generosity, both while I was a visitor at Pittsburgh and continuing afterwards, has been critical in the shaping of this project. Jim's work, and conversations with Jim, have been no less influential to the shaping of the project, even though Jim's name appears just a few times in the text. It would not be an exaggeration to say that much of my basic conceptual repertoire, the class of concepts with which I now do philosophy, is owed to countless hours of working with and talking to Jim.

Apart from my committee, I've benefited from conversations with many great members of the faculty at Chicago, especially David Finkelstein, Matt Boyle, Irad Kimhi, Chris Kennedy, Kevin Davey, Matthias Haase, Ginger Schultheis, and Ben Callard. There are many (now mostly former) graduate students (and visitors) at Chicago I'd like to

thank for stimulating and challenging conversations. Most of all, I am grateful for very many conversations with Lawrence (Dusty) Dallman which have greatly shaped my conception of the very topic of this dissertation. I am also extremely grateful for many conversations with Till Hoeppner as well as Nic Koziolek, Matt Teichman, Andrew Pitel, Martijn Wallage, Patrick Muñoz, Julian Grove, Dries Daniels, Joe Brewer, Molly Brown, and many other graduate students and visitors at Chicago. Also, many thanks to William Weaver for unwavering administrative support throughout the years which made my life as a graduate student at Chicago immeasurably easier. Outside of Chicago, there are many people who I've had the benefit of discussing these topics with, especially Quentin Fischer, Aaron Salomon, Douglas Vaaler, Tom Breed, Eric Marcus, Zach Wrublewski, Jason Lemmon, Mack Sullivan, and Alex Rausch (who gets a special thanks for waking me out of my dogmatic slumber my first year in graduate school). Several specific conversations have had an impact on my thinking over the years, particularly conversations with Jeff King my first year as a graduate student, Peter Hanks my third year, and, recently, Matt Mandelkern. Finally, I'd like to thank Quill Kukla for getting me interested in this sort of stuff as an undergraduate in the first place.

Much of the work here, especially in the positive part, was the result of collaboration with the Research on Logical Expressivism (ROLE) working group, led by Bob Brandom and Ulf Hlobil, whose members also currently include Dan Kaplan, Shuhei Shimamura, Rea Golan, and Viviane Fairbank. I'd like to thank all the members of that group—especially Ulf and Dan (and, of course, Bob again)—for their helpful engagement with various half-baked ideas of mine, many of which eventually found their way, in some form, into the dissertation. My thinking on Sellars has also been informed by participation in the International Sellars Colloquium (ISC) over the past few years. I'd like to thank Luz C. Seiberth for inviting me to the group, and I am thankful for comments on my work on Sellars from members of the group, especially Bill DeVries, Jim O'Shea, Lionel Shapiro, Michael Hicks, Preston Stovall, and Robert Kraut.

On a personal note, endless thanks to Haley Church for all her support through the final stretch of this project. Finally, a special thanks to my mom for editing each and every chapter of this dissertation. I hope it wasn't too boring.

Introduction

In March 2006, a post by Jason Stanley appeared on the Leiter Report blog entitled "The Use Theory of Meaning." The post purported to be a sounding of the death knell for semantic theories that take the notion of use, of cognitive or linguistic role, rather than the notions of reference or truth, as primitive. Such theories, so called "conceptual role semantics," are widely regarded, at least by those sympathetic to them, as "the main rival to theories that take notions such as truth or reference as central," (Whiting 2006). However, from the "orthodox" perspective occupied by Stanley, there is no real rivalry at all. If you look at the class of people in philosophy and linguistics who call themselves "semanticists" a very tiny subset of these people are doing conceptual role semantics. Nearly every semanticist works in a style of semantic theory that takes notions such as truth or reference as central. Among this orthodoxy, it seems that there is fruitful debate and real progress concerning particular proposals for various classes of expressions against the backdrop of general agreement on the framework in which semantic questions are to be answered. On the other hand, while several philosophers have proposed conceptual role semantics as a preferable alternative to truth- or reference-based semantic theories, there are no agreed upon reasons for this preference nor is there any agreed upon framework in which conceptual role semantics can actually be done. It is from this perspective that Stanley, speaking on behalf of theorists who "operate with the notions of reference and truth" says, of those operating fundamentally with the notion of cognitive or linguistic role, "we regard their work at best as useless for the philosophical project of understanding the language-world relation, and at worst as a vain attempt to reinvent the wheel," (Stanley 2006).

Among the targets of Stanley's criticism is the type of the semantic theory put forth schematically in the work of Wilfrid Sellars (1953, 1954, 1974) and articulated in detail in Robert Brandom's (1994) *Making It Explicit* in which the meaning of an expression is

understood in terms of the norms governing its use in discourse.[1] Stanley, and apparently most contemporary semanticists with him, regards this type of theory as "useless for the philosophical project of understanding the language-world relation." Now, if one thinks of such a semantic theory in some of the terms in which Brandom has put it, for instance, as eschewing "word-world" relations in favor of "word-word" relations (Brandom 1984), this claim of Stanley's will not be too surprising. From a Sellarsian perspective, however, Stanley's claim is striking. Sellars takes it that *only* a version of conceptual role semantics enables us to understand the relation between language and the world. It is my aim in this dissertation to substantiate and defend this Sellarsian claim. Like any claim involving the use of the word "only," this Sellarsian claim has both a negative component, ruling out, and a positive component, ruling in. Accordingly, this dissertation has both a negative part and a positive part.

Negatively, I'll argue that truth-conditional semantics, which I'll articulate as a species of what I call "worldly semantics," is not able to provide us with an understanding of the relationship between language and the world. A worldly semantic theory is a theory that takes knowledge of meanings to be asymmetrically dependent on knowledge of worldly entities and their relations. Such "worldly entities" could be possible worlds, or they could be such things as objects and properties in the actual world. Any semantic theory that takes knowledge of semantic facts, such as the fact that the sentence "*a* is gray" is incompatible with the sentence "*a* is white," to be asymmetrically dependent on knowledge of worldly facts, such as the fact that the set of possible worlds in which *a* is gray is disjoint from the set of possible worlds in which *a* is white or the fact that the property of being gray and the property of being white cannot be jointly instantiated by some object, is a version of worldly semantics. I divide worldly semantic theories into two main varieties, which I call "extra-worldly" semantics and "intra-worldly" semantics. These are the targets of

[1]Though Brandom's theory is not mentioned in the body of the post, it is clearly a target. Brandom describes *Making It Explicit* as " an attempt to explain the meanings of linguistic expressions in terms of their use," (1997, 153), and it explicitly comes under attack by Stanley in the comments, particularly in an exchange with Mark Lance, who defends a variant of this theory. The theory developed there has been called "normative inferentialism" (Lance 1996, Peregrin 2014), "normative functionalism" (Maher 2012), "normative dynamics," (Nickel 2013), and various other names. I add to the list, calling it "discursive role semantics" here, but I am not particularly attached to that name (in part because, in certain formal contexts, abbreviating it would lead to confusions with the "discourse representation structures" (DRSs) of Kamp's (1981) discourse representation theory).

chapters two and three, respectively. I argue that both variants of worldly semantics fall prey to what Sellars (1956) calls "the Myth of the Given." Though this term is often thrown around, there is no general agreement on either what it picks out or what the problem with a theory that is picked out by it is. In the first chapter of the dissertation, I will say just what it is for a philosophical conception to be an instance of the Myth. In the next two, I will show how both the extra- and intra-worldly variants of worldly semantics are such instances, and that their being such really is fatal to these worldly semantic theories, at least insofar as they aspire to *account for* or *explain* our knowledge of meaning, rather than simply to *elucidate* or *explicate* it.

Positively, I will argue that, unlike worldly semantics, the species of conceptual role semantics put forward by Sellars and developed by Brandom, which I'll call "discursive role semantics," is able to provide us with an understanding of the relationship between language and the world. The key idea involves an inversion of the order of explanation presupposed by worldly semantics. Rather than taking our semantic knowledge to be asymmetrically dependent on worldly knowledge, it is argued that what worldly semantic theories take to be worldly knowledge is nothing other than our semantic knowledge, articulated in a worldly mode. Though this claim has been made by Sellars and Brandom, it has not been developed in the context of a formal semantic framework. I'll do that here. Crucially, on a discursive role semantic theory, the semantic values assigned to expressions of that language are not dependent on a pre-given domain of extra-linguistic entities. Rather, semantic values are articulated entirely in terms of the rules governing the use of expressions in the language. Accordingly, knowledge of meaning is not taken to asymmetrically depend on worldly knowledge, but, rather, is understood in entirely intra-linguistic terms. This conception of the semantics may prompt worries of linguistic idealism, but, in the final chapter, I argue that, on the contrary, only such a semantic theory can avoid the problematic idealism that is implicit in worldly semantics. Discursive role semantics enables us to draw a distinction between the "world" of conceptual contents conferred by a certain linguistic practice and the real world to which that practice really belongs. Once this distinction is in view, discursive role semantics enables us to make sense of the real relation between language and the world.

Throughout much of this dissertation, I will be making use of a very simple toy language, meant to encode a minimal bit of semantic content. With the use of this toy language, one can say that something is *white*, *gray*, or *black*, that something is *darker than*, *lighter than*, or *the same shade as* something else, and that something is *not* the case, that something *and* something else is the case, and that something *or* something else is the case. The point of introducing such a simple toy language is to be able to get the entirety of a set of semantic theories for the same language easily in view, so that their overall structure can be examined side by side. I went back and forth at various stages in the writing process between introducing a more complex toy language and going back to this very simple one. I ended up sticking with the simple one that is contained here, content with the conclusion, which you will have to verify for yourself, that the introduction of a more complex toy language would have only obscured the basic point I hope to demonstrate. The basic point I hope to demonstrate with the use of this very simple toy language in both the negative and positive part of this dissertation is that there is a fundamental problem with worldly semantics, of both the extra- and intra-worldly variety, that discursive role semantics resolves. In the positive part of the dissertation, I expand the toy language to include quantified strict conditionals so that its speakers can say such things as "Necessarily, if something's black, then it's darker than anything gray." However, the point of introducing this additional vocabulary is just to illustrate the basic philosophical proposal. The project of actually carrying out discursive role semantics for natural language is left for other work. My aim in the present work is to motivate this project against the currently dominant worldly semantic paradigm on philosophical, rather than empirical, grounds.

This dissertation is divided into six chapters: three negative and three positive. In Chapter One, "Worldly Semantics and the Myth of the Given," I lay out the aim of an explanatory semantic theory, the basic structure of the genus of semantic theory that I call "worldly semantics" and the form of the Mythical conception of the relation between mind and world to which any worldly semantic theory is committed. In Chapter Two, "Exrta-Worldly Semantics," I will lay out the species of worldly semantics that I call "extra-worldly semantics," whose principle philosophical advocates are David Lewis (1973, 1986)

and Robert Stalnaker, and argue that it suffers from a fatal instance of the Myth of the Given. In Chapter Three, "Intra-Worldly Semantics," I consider a different version of worldly semantics whose principle advocates include, among others, Scott Soames (2010, 2014, 2015) and Jeff King (2007b, 2014), and argue that it too (albeit in a different way) suffers from a fatal instance of the Myth of the Given. In Chapter Four, "Discursive Role Semantics," I spell out a version of the alternative, non-worldly semantic theory that I endorse, which I call "discursive role semantics," whose principle advocates are Sellars (1953, 1954, 1974), and Brandom (1994). In Chapter Five, " 'Worldly' Knowledge as Semantic Knowledge," I expand the toy language to include quantifiers and modal operators drawing on technical work by Mark Lance and Philip Kremer (1994), with the basic aim of spelling out what Amie Thomason (2020) calls a "modal normativist" conception of "worldly" knowledge appealed to in worldly semantic theories, where this knowledge is conceived as really semantic knowledge, expressed in a worldly mode. In Chapter Six, "Language and the World," I develop the alternate conception of the relation between language and the extra-linguistic world afforded by discursive role semantics, offering an integrated conception of semantics in a scientific worldview.

1

Worldly Semantics and the Myth of the Given

1.1 Introduction

In this opening chapter, I'll lay out the explanatory aim of a semantic theory to be, and I introduce "worldly semantics" as a strategy for accomplishing this aim. Defining a simple toy language to use as our example, I'll explicate the basic structure of such a theory. I'll then state the thesis, which I'll defend in the next two chapters, that worldly semantics is committed to an instance of what Wilfrid Sellars (1956) calls "the Myth of the Given," spelling out at some length what I take this term to pick out. Finally, I'll clarify the target by drawing a distinction between "explanatory" and "elucidatory" models in semantics and illustrating what is at stake in locating, for instance, possible worlds semantics on one side of that distinction rather than the other.

1.2 Our Semantic Aim

Language speakers aren't mindless automota. Generally, they know what they're saying when they use expressions of a language that they know how to speak. They have this knowledge because they know what these expressions mean. The aim of semantics is to understand what it is in which this aspect of the capacity to speak a language, knowledge of meaning, consists. Here is one particularly clear statement of this aim from Seth Yalcin (2018):

> I take it that in natural language semantics, the aspect of reality we are seeking some understanding of is a dimension of human linguistic competence—informally, knowledge of meaning. Competent speakers of a language know ('cognize', etc.) the meaning features of expressions of their language. The

> semanticist is interested in modeling this state of mind and the associated semantic features, (2018, 353).

I take it that most semanticists working in the Chomskian tradition of generative grammar think of the discipline roughly along these lines.[1] In semantics, we are aiming to understand the knowledge of meaning that competent speakers have. We take this knowledge to explain certain aspects of their linguistic behavior—the behavior that they exhibit in virtue of knowing what expressions of their language mean—and what we're aiming to do, in constructing a semantic theory, is to explain this behavior by modeling the knowledge of meaning that accounts for it. The main task of the semantic theorist is to assign *semantic values* to the expressions of the language for which she is constructing a semantic theory. These semantic values are the entities in the semantic theory that are meant to be mathematically defined models of the meanings of the expressions of which speakers of that language have knowledge. So, meanings are theorized to play a certain sort of explanatory role: knowledge of them is taken to explain certain aspects of speakers' behavior. And semantic values are entities in the theorist's model that are mathematically defined in such a way that they satisfy certain properties, properties which are either identical or structurally analogous to the properties that the theorist takes it that meanings must have in order for them to play the explanatory role that they are theorized to play.

Now, there are several properties that the meanings of sentences are taken to have that are meant to be modeled by the assignment of semantic values to them. To limit the scope of my discussion, I will focus here on just one crucial such property. The meanings of sentences are taken to determine facts consisting in these sentences standing in relations of entailment and (in)compatibility (or, consequence and (in)consistency) to one another. One important role of semantic values is to model meanings in such a way that we can explain these facts, thereby explaining speakers' knowledge of them, and thereby explaining their behavior that is a manifestation of this knowledge. Here is Yalcin again, in a different paper of his, stating this point:

[1]For instance Gennaro Chierchia and Sally McConnell-Ginet (1990) write "It is the application of mathematical models to the study of the cognitive phenomenon of linguistic knowledge that most generative linguists recognize as their aim," (2). For an overview of the Chomskyian contextualization of the formal methods developed by the philosophical pioneers of semantic theorizing in contemporary semantics, see Soames (2019, 133-156).

> [S]emantic values are assumed to be the sorts of things consequence and consistency relations are articulated in terms of: when $\Gamma \vdash \varphi$ holds, this is (at least partly) because of the semantic values of (the sentences in) Γ and of φ, respectively. Hypotheses about semantic values can thereby serve to predict, and ground, entailment and consistency facts, hence knowledge of such facts, (2014, 24).

Assigning semantic values to sentences, and then articulating consequence and consistency relations in terms of these semantic values, we model the meanings of sentences in such a way that we are able to explain, with the use of our model, how facts consisting in sentences entailing or being incompatible with one another obtain in virtue of these sentences meaning what they do. Then, by modeling of speakers' knowledge of the meanings of these sentences as knowledge of the semantic values we assign to them, our theory will explain how their knowledge of entailment and incompatibility relations obtaining between sentences, knowledge which explains certain aspects of their behavior, is determined by their knowledge of the meanings of these sentences.

1.3 A Toy Language

To see more determinately what aspects of speakers' behavior we're trying to explain in assigning semantic values to sentences in terms of which entailment and incompatibility relations can be articulated, it will be helpful to introduce a very simple "toy language" and then consider what theoretical work a semantic theory for this toy language should be able to do insofar as it aspires to this explanatory aim. So, imagine a small linguistic community whose members speak a language consisting of the following expressions:

1. Three names: "*a*," "*b*," and "*c*"
2. Three 1-place predicates: "is white," "is gray," and "is black"
3. Three 2-place predicates: "is lighter than," "is darker than," and "is the same shade as"
4. One unary sentential operator: "It is not the case that"
5. Two binary sentential operators: "and" and "or"
6. Left and right parentheses (to avoid ambiguity): "(" and ")"

This is their basic vocabulary: the set of simple expressions that they are able to employ. The grammar of their language, through which complex expressions can be constructed from these simple ones, can be recursively specified as follows:

1. Any name followed by a 1-place predicate is a sentence.
2. Any name followed by a 2-place predicate and then another name is a sentence.
3. If φ is a sentence and U is a unary operator ($U\varphi$) is a sentence.
4. If φ and ψ are sentences and B is a binary operator ($\varphi B\psi$) is a sentence.
5. If some string of lexical items can't be constructed by the use of these rules, it's not a sentence.

Call any sentence that contains no sentential operators an "atomic sentence." There are thirty-six atomic sentences of our toy language, including, for instance "a is white," "b is darker than c," "c is gray," and so on. There are infinite non-atomic sentences, formed by conjoining atomic sentences with operators and parentheses. Our toy language consists in this infinite set of sentences. This is, of course, a woefully impoverished language, and it can hardly be called a language at all, but it is enough of a language for our purposes here.

I will make extensive use of this toy language throughout this dissertation, so it is worth saying a few words now to preliminarily justify my doing so. As it is probably clear, I have laid out the simplest toy language that I possibly could. I hope it will be clear in what follows that I could have introduced a much fancier toy language here—with more vocabulary belonging to the grammatical types introduced here, vocabulary belonging to additional grammatical types, and a more complex grammar to accommodate this additional vocabulary—and used it to the same end in the next few chapters. I take it that doing this would have only unnecessarily complicated things, so that is why I have not done so, saving the introduction of more sophisticated toy languages to the positive part of the dissertation where I develop the alternative framework of discursive role semantics.[2] Of course, it takes some cognitive dissonance to imagine that we could

[2]It is also worth pointing out that, grammatically, I have taken some shortcuts and treated this toy language in such a way that much more closely resembles the formal language of first-order logic than a natural language like English. Once again, this is just for simplicity, and nothing important hangs on this. Contemporary work in semantics, following Montague (1974), "reject[s] the contention that an important theoretical difference exists between formal and natural languages," (188).

really have what Brandom calls an "autonomous discursive practice" whose members speak this "language" and it alone. Indeed, as we'll later see, there could not be such a practice, and we will need a richer language in order to be able to think of there as being speakers who employ that language and it alone. For now, however, let us engage in the imaginative exercise of taking there to be speakers who speak this language and it alone, grasping the meanings of the expressions that belong to it and behaving certain ways in virtue of grasping these meanings.

Suppose our speakers act in such a way that shows that they take the sentences "*a* is darker than *b*" and "*b* is lighter than *a*" to be synonymous. Now, if they have semantic vocabulary, they might say "These two sentences are synonymous," "These two sentences mean the same thing," or "One says the same thing in uttering either of these two sentences," but we need not even credit them with this sort of vocabulary in order to get our basic explanandum into view; it is sufficient that they, in their linguistic practices, treat the two sentences in the way that two synonymous sentences ought to be treated. For instance, whenever a competent speaker utters one, they'll be prepared to utter the other, if an incompetent speaker utters one but refuses to utter the other, they'll be corrected by competent speakers, and so on. Similarly, competent speakers of this language take the sentences "*a* is black" and "*b* is gray" to jointly entail the sentence "*a* is darker than *b*." Any competent speaker that utters both "*a* is black" and "*b* is gray" will also be prepared to utter "*a* is darker than *b*," and if an incompetent speaker utters the first two but refuses to utter the third, they'll be corrected, and so on. Finally, they take the sentences "*a* is gray" and "*a* is white" to be incompatible. They'll never utter both sentences at the same time, they'll correct incompetent speakers that do, and so on. These activities, we theorize, are manifestations of their knowledge of the meanings of the sentences "*a* is darker than *b*," "*b* is lighter than *a*," "*a* is black" "*b* is gray," "*a* is gray," and "*a* is white." That is to say, it is in virtue of knowing what these sentences mean that the speakers of our toy language behave in these ways. What we want to do, in constructing a semantic theory for their language, is understand this knowledge of meaning in such a way that enables us to explain this behavior. Officially, what we want to do is assign semantic values to these sentences, formal models of their meanings, such that if speakers know that these

sentences have these semantic values, they'll know these sentences stand in these semantic relations, since, if they know that these sentences stand in these semantic relations, they'll behave in these ways.

Given this explanatory aim, assigning semantic values to sentences should enable us to account for facts like the following:

F1. The sentence "*a* is darker than *b*" is synonymous with the sentence "*b* is lighter than *a*."

F2. The sentences "*a* is black" and "*b* is gray" jointly entail the sentence "*a* is darker than *b*."

F3. The sentence "*a* is gray" is incompatible with the sentence "*a* is white."

If, by assigning semantic values to the sentences "*a* is darker than *b*," "*b* is lighter than *a*," "*a* is black" "*b* is gray," "*a* is gray," and "*a* is white," we are able to account for these facts, then, by thinking of speakers' knowledge of the meaning of these sentences in terms of their knowledge of these semantic values, we can explain their knowledge of these facts, and, accordingly, the behavior they exhibit in virtue of having this knowledge.

1.4 The Meanings of Content Words

In specifying (F1)-(F3), I have picked out by way of example only one class of synonymy, entailment, and incompatibility relations that obtain between sentences of this toy language: the class of *material* rather than *formal* relations of synonymy, entailment, and incompatibility. For instance, the sentences "*a* is gray" and "*a* is white" are *materially* incompatible, whereas the sentences "*a* is gray" and "It's not the case that *a* is gray" are *formally* incompatible. Articulating this distinction with the use of a more contemporary vocabulary, material semantic relations are relations that obtain between sentences in virtue of the meanings of (what are often called) the "content words" contained in those sentences, words like "gray," "white," or "darker than," whereas formal semantic relations are relations that obtain between sentences in virtue of the meanings of (what are often called) the "function words" like "not," "and," and "or."[3] Now, articulating exactly

[3]See, for instance, Lobner (2002, 4-5) and Szabó (2019) for an articulation of this distinction with the use of this terminology. This terminology is somewhat confusing, since the meanings of content words are

what this distinction consists in will depend on the sort of semantic theory that one ends up endorsing. In almost every semantic theory, however, the semantic values assigned to content words will form the foundation on the basis of which the rest of theory will be constructed.[4]

It is only given the assignment of semantic values to these simple content words that the assignment of semantic values to function words, generally conceived in terms of operations on the semantic values of content words, makes any sense at all. For instance, in a possible worlds semantics, semantic values for logically complex sentences can be understood in terms of the set-theoretic operations of complementation, intersection, and union only insofar as atomic sentences are assigned sets of possible worlds as semantic values, and atomic sentences can be assigned sets of possible words as semantic values only insofar as the content words that make them up, words like "gray" or "white," are assigned suitable semantic values, for instance, functions that map each possible world to the set of things that are gray in that world or white in that world. So the assignment of suitable semantic values to content words is required at the base level of semantic theories. This fact about the structure of semantic theories follows directly from such a theory's commitment to the compositionality of meaning: that the meaning of a complex sentence is determined by the meaning of its parts and the way those parts are put together. If we cannot think of the meanings of content words as adequately modeled by the semantic values that a compositional semantic theory assigns to them, the whole theory that is based on these basic assignments falls like a house of cards.

Despite the fact that semantic theories require the assignment of semantic values to content words at their base level, most semanticists do not concern themselves with these basic assignments of semantic values. While it does fall to the semantic theorist to specify the semantic types corresponding to content words of different syntactic categories, the task of specifying the meanings of these basic expressions in any substantive way is not a task for semantics, properly construed. Distinctions in meaning between such words

themselves taken to be functions. Kearns (2011) uses the terms "categorematic" and "syncategorematic," but this is also potentially problematic, given that the meanings of those terms, in a contemporary context, don't directly map on to their classical usage.

[4]This is, I take it, implicitly acknowledged by almost all semantic theorists; one theorist who is explicit about this is Szabó (2019).

as "gray" and "white," insofar as they belong to the same syntactic category, are, "from the point of view of semantic theory, simply brute," (Yalcin 2018, 350). So, for instance, a possible worlds semantics might assign to the 1-place predicate "gray" the function that maps each world to the set of things that are gray in that world, and it will assign "white" the function that maps each world to the set of things that are gray in that world. Of course, such a theory won't tell us what it is for something to be gray as opposed to white, but we shouldn't expect it to. A dictionary can tell us this. To think that it's the job of the semantic theory to tell us what a dictionary would tell us would be to confuse semantics with lexicography, and that, as Richard Thomason (1975) says, is "a persistent and harmful source of misunderstanding in matters of semantic methodology," (48). Yalcin quotes this sentiment, expressed by Thomason in his introduction to Montague's *Formal Philosophy*, in support of this attitude towards the meanings of content words:

> *[W]e should not expect a semantic theory to furnish an account of how any two expressions belonging to the same syntactic category differ in meaning* . . . 'Walk' and 'run', for instance, and 'unicorn' and 'zebra' certainly do differ in meaning, and we require a dictionary of English to tell us how. But the making of a dictionary demands considerable knowledge of the world [of a sort the semantic theorist should not be expected to furnish]," (Yalcin 2018, 350, quoting Thomason (1975); Thomason's italics, Yalcin's bracketed addition.)

Knowing the meanings of content words like "walk" and "run" or "gray" and "white" requires "considerable knowledge of the world." A semantic theorist, in taking speakers to have knowledge of the meanings of words like "walk" and "run" or "gray" and "white," appeals to this worldly knowledge that speakers have—their knowledge of what it is for something to walk as opposed to run, or what it is for something to be gray as opposed to white—but this worldly knowledge, which is an ingredient in speakers' knowledge of meaning, is to be distinguished from the properly semantic knowledge that is the proper concern of the semantic theorist. As such, it is sufficient for the semanticist to say something along the following lines:

> The meaning of the predicate "gray" is determined entirely (or, at least, sufficiently for our purposes) by the fact that it is correctly applied to some object just in case that object is gray. Accordingly, we can model the meaning of the predicate "gray" as a function that maps each possible world, each way for

> things to be, to the set of things that are gray in that world; the set of things to which that predicate is correctly applied. What we model in modeling the meaning of this predicate in this way is what a speaker knows in knowing the meaning of this predicate; they know that, however things are, this expression is to be applied to something just in case that thing is gray.

Having given this justification of their formal model of meanings of 1-place predicates such that they compose in the right ways with the meanings of other types of expressions, the semanticist can leave it to the lexicographer to say, substantively, what it is for something to be gray, how being gray differs from being white, and so on.

This apparent division of labor may seem to be of a piece with the divide-and-conquer methodology found throughout the natural sciences. However, implicit in this way of thinking about speakers' knowledge of meaning is the idea that speakers' knowledge of certain worldly facts—for instance, the fact that the property of being gray is incompatible with the property of being white (i.e. being gray is a way for something to be such that, if something is that way, it cannot be white)—is explanatorily prior to their knowledge of certain semantic ones—for instance, the fact that the predicate "gray" is incompatible with the predicate "white." It is because the former sort of knowledge, the worldly knowledge, is not taken to be the proper object of a semantic theory that the knowledge of the incompatibility of the predicates, from the point of view of the semantic theory, can be taken to come for free as a direct consequence of semantic values for content words that are "from the point of view of semantic theory, simply brute," (Yalcin 2018, 350). For instance, in a possible worlds semantics, one simply assigns the predicates intensions that are assumed to be disjoint, appealing to one's own knowledge of what it is for something to be gray or white in the assessment of these intensions as disjoint. This appeal to one's own knowledge of what it is for something to be gray or white can be taken to be unproblematic only insofar as this knowledge that one appeals to is taken to be "knowledge of the world of a sort the semantic theorist should not be expected to furnish," (Yalcin 2018, 350). This worldly knowledge is taken to underlie the semantic knowledge that constitutes of the base of the semantic theory, the knowledge of the meanings of content words like "gray" and "white" that grounds knowledge of facts such as (F1)-(F3). The semantic theories I will concern myself with in the negative part of this dissertation all assign semantic values

to content words in accord with this theoretical orientation. They are all versions of what I will call "worldly semantics." Let me now lay out, in abstract terms, the basic structure of a worldly semantic theory.

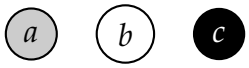
1.5 The Basic Structure of Worldly Semantics

Most work in contemporary semantics is guided by the following core idea, which I'll quote directly from the introductory textbook in formal semantics by Dowty, Wall, and Peters (1981):

> To know the meaning of a (declarative) sentence is to know what the world would have to be like for the sentence to be true, (4).[5]

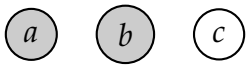
To see how this core idea applies to the speakers of our toy language, let's suppose that there are only three things in the world in which they live—*a*, *b*, and *c*—and that these are the things that are named by the names "a," "b," and "c." Furthermore, let's suppose that there are only three ways that these three things can be—completely white, completely gray, or completely black—and that these are the ways the speakers of our toy language say that something is when they say of it that it "is white," "is gray," or "is black." Finally, let's suppose that there is only one shade of gray, so if something is darker or lighter than something else, it can't be the case that they're both gray.

Now, consider the sentence "*a* is gray." There are many ways that the world of the speakers of our toy language can be such that this sentence is true. The world can be such that *a* is gray, *b* is white, and *c* is black. That is, the world can be like this:



Alternately, the world can be such that *a* is gray, *b* is gray, and *c* is white. That is, the world can be like this:

[5] See also the widely used introductory textbook by Heim and Kratzer (1998), where they open with the sentence,"To know the meaning of is to know its truth conditions," going on to tell us, to know the meaning of a sentence, you don't have to know whether it is true; "What you do know, however, is what the world would have to be like for it to be true," (1).



As far as the truth of "*a* is gray" is concerned, whether *b* is white or *b* is gray or whether *c* is black or *c* is white does not matter. What does matter is whether or not *a* is gray. The sentence "*a* is gray" is true just in case the world is such that *a* is gray. To know the meaning of this sentence, on a truth-conditional theory of the sort described by Dowty, Wall, and Peters, is to know just this.

Now, there are different ways in which this basic idea, expressed by Dowty, Wall, and Peters, can be implemented in a semantic theory in order to arrive at formally specificiable semantic values that are meant to serve as models of what speakers know in knowing the meaning of a sentence. I will consider what I take to be the two basic ways in which this idea can be implemented in the next two chapters. For now, however, let's consider the general structure of a theory that conforms to this basic idea. Such a theory will be a version of what I will call "worldly" semantics. On a worldly semantic theory, we take speakers' knowledge of meaning to asymmetrically depend on their knowledge of "worldly" entities and their relations. The general sorts of "worldly" entities to which a worldly semantics appeals might be picked out with expressions such as "objects," "properties," "relations," "states of affairs," "possible worlds," and so on. Which of these sorts of entities are given priority over the others will vary from theory to theory, but, in a worldly semantics, speakers are taken to have knowledge of entities of these sorts and knowledge of relations that entities of these sorts stand to one another, and their knowledge of the meanings of sentences of their language is taken to depend on this worldly knowledge. To see how a worldly semantics is supposed to work, consider the sort of explanation of (F3) specified above—the fact that the sentence "*a* is gray" is incompatible with the sentence "*a* is white"—that we would provide if we endorse a worldly semantics.

We start by taking the speakers of our toy language to have a bit of worldly knowledge: They know that if *a* is gray, then it can't be the case that *a* is white. This bit of worldly knowledge might be analyzed in different ways. We might analyze it in terms of ways the world as a whole can be, saying that speakers know that, however the world can possibly

be, the set of things that are gray in the world and the set of things that are white in the world are disjoint—they don't have any elements in common. Spelling this out a bit, we might take our speakers to have a grip on a space of different possible ways for the world as a whole to be, a space of different "possible worlds," and know that, for each point in this space, each possible world, the set of things that are gray and the set of things that are white are disjoint—these two sets do not have any common elements. So, whichever element of the total set of possible worlds is actual, if *a* is an element of the set of gray things, then it isn't an element of the set of white things. This is one way to analyze what it is that our speakers know in knowing that, if *a* is gray, then it can't be the case that *a* is white. Since the worldly entities of which speakers are taken to have knowledge are worlds as a whole, I'll call it an "extra-worldly" analysis. Alternately, we might analyze our speakers' knowledge of the fact that if *a* is gray, then it can't the case that *a* is white entirely in terms of ways things in this world can be. We might say that our speakers know that no single thing can, at the same time, be both gray and white. Spelling this out a bit, we might say that there is a basic modal fact about the property of being gray and the property of being white, one that obtains in virtue of the essences of these two properties; it is not possible for a single thing to, at one time, instantiate both the property being gray and the property of being white. So, since *a* is a single thing, if *a* instantiates the property of being gray, it is not possible for it, at the same time, to instantiate the property of being white. This is a different way to analyze what it is that our speakers know in knowing that, if *a* is gray, then it can't be the case that *a* is white. Since the basic worldly entities of which speakers are taken to have knowledge here are things *in* the world, objects and properties that these objects might have, I'll call it an "intra-worldly" analysis. However we want to analyze our speakers' worldly knowledge of the fact that if *a* is gray, it can't be the case that *a* is white, if we endorse a worldly semantics, we'll think of our speakers' knowledge of the fact that the sentence "*a* is gray" is incompatible with "*a* is white" as asymmetrically depending on a bit of worldly knowledge of this sort.

Consider first the semantic picture suggested by an extra-worldly conception of this bit of worldly knowledge. On a standard variant of extra-worldly semantics, the semantic value of an expression is a function that maps possible worlds to extensions. For names,

these extensions are taken to be particular objects, and for 1-place predicates, these extensions are taken to be sets of objects. So, the semantic value of "*a*" is a function that maps each possible world to *a*, a particular thing that we may assume exists in each world, and the semantic value of "is gray" is a function that maps each possible world to the set of things that are gray in that world.[6] Now, we have a rule of composition that says that, for any sentence of the form "*n* is *F*," consisting in a name "*n*" concatenated with a 1-place predicate "is *F*," the semantic value of that sentence is the set of worlds *w* such that the object to which the semantic value of "*n*" maps *w* is an element of the set of objects to which the semantic value of "*F*" maps *w*. So, the semantic value of "*a* is gray" is the set of worlds in which *a* is an element of the set of gray things in that world. Likewise, the semantic value of "*a* is white' is the set of worlds in which *a* is an element of the set of white things in that world. Now, if one knows that, for each possible world, there is no object that is an element of both the set of gray things in that world and the set of white things in that world, and one knows that "*a* is gray" and "*a* is white" have the semantic values that they do, one will know that the sets of worlds that are the semantic values of "*a* is gray" and "*a* is white" are disjoint. That, according to an extra-worldly semantics, is just what it is to know that the sentences "*a* is gray" and "*a* is white" are incompatible. In this way, an extra-worldly semantics takes speakers' semantic knowledge to asymmetrically depend on a bit of extra-worldly knowledge.

Now consider the semantic picture suggested by an *intra*-worldly conception of this bit of worldly knowledge. On a standard variant of an intra-worldly semantics, the semantic value of a name is the object named by that name, and the semantic value of a 1-place predicate is the property expressed by that predicate. So, the semantic value of "*a*" is *a*, the object that is named by "*a*," and the semantic value of "gray" is the property of being gray, the property that is expressed by "gray."[7] Now, we have a rule of composition that

[6]This assumes, following Kripke (1980), that names are "rigid designators." At some point, we'd have reason to drop the assumption that *a* exists in each world. In which case, we can take the semantic value of a name to be a *partial* function that maps each possible world in which the object actually named by that name exists to that object. I'll continue to make such simplifying assumptions here.

[7]This will need some refinement, depending on the variant of intra-worldly semantics we consider. Soames (2014, 2015), for instance, takes the semantic values of names and predicates not to be objects and properties themselves, but acts of cognizing objects and properties. Once again, I'll make simplifying assumptions in the consideration of intra-worldly semantics, but nothing will hang on this.

says that, for any sentence of the form "*n* is *F*," consisting in a name "*n*" concatenated with a 1-place predicate "is *F*," the semantic value of that sentence is a structured proposition that represents the object that is the semantic value of "*n*" as instantiating the property that is the semantic value of "*F*." So the semantic value of "*a* is gray" is a structured proposition that represents *a* as instantiating the property of being gray. Likewise, the semantic value of "*a* is white" is a structured proposition that represents *a* as instantiating the property of being white. Now, if one knows that the property of being gray and the property are being white are incompatible in the sense that it is not possible for a single object to instantiate both of these properties, one will know that these two propositions that are the semantic values of "*a* is gray" and "*a* is white" cannot both be true; taken together, they represent a single thing as being two ways that a single thing cannot, at a single time, be. That, according to an intra-worldly semantics of this sort, is just what it is to know that the sentences "*a* is gray" and "*a* is white" are incompatible. In this way, an intra-worldly semantics takes speakers' semantic knowledge to asymmetrically depend on a bit of intra-worldly knowledge.

Though this theoretical structure is rarely as explicit as I have made it out here, I take it that worldly semantics is pretty much ubiquitous in contemporary theorizing about speakers' knowledge of meaning in both philosophy and linguistics. Almost every working semanticist practices a variant of worldly semantics. Semantic theories that are often taken to be on opposite sides of fundamental dividing lines in semantic theorizing—for instance, truth-conditional semantics vs. dynamic semantics, possible worlds semantics vs. situation semantics—will generally still all be variants of worldly semantics.[8] Nevertheless, I will argue here that such semantic theories are not able to do the explanatory work that a semantic theory is supposed to be able to do. They are not able to give

[8]It's worth noting that, though dynamic semantics of the sort proposed by Veltman (1996) do not conform to the basic truth-conditional dictum quoted above from Dowty, Wall, and Peters, they nevertheless require conceiving of extra-worldly knowledge as underlying knowledge of meaning. I'll address such theories explicitly in Chapter Four. Other than the small minority of proponents of inferential role semantics mentioned in the introduction, there are some notable exceptions to the worldly semantic orientation in contemporary semantics. Perhaps most notably, there is Paul Pietroski's (2018) Chomskyian internalism, according to which meanings are instructions for conceptual operations. I leave open the question of whether the criticisms of worldly semantics developed here apply to this approach and, to whatever extent that they do not, how the relationship between this approach and the one developed here in the positive part of the dissertation should be understood.

us anything resembling an account of the aspect of semantic competence consisting in knowledge of meaning. The basic problem is this: we cannot give an account of speakers' knowledge of meaning as asymmetrically depending on their knowledge of worldly entities and their relations in the way that a worldly semantics requires us to do because this worldly knowledge itself depends on speakers' knowledge of meaning. To give a name to the general form of the problem of which this problem is a specific instance and to give some philosophical context for the line of critique I am about to prosecute, my claim is that, in proposing to explain speakers' knowledge of meaning in the way that they do, proponents of worldly semantics are guilty of a version of what Wilfrid Sellars calls "The Myth of the Given," (1956).

Now, my main aim in this dissertation is not the exegesis of Sellars.[9] I do take it, however, that all of my claims, both negative and positive, are basically Sellarsian ones. Accordingly, it is worth doing a little work to articulate, in Sellars's own terms, what the general form of the problem is and how, by his own lights, worldly semantics is an incarnation of it.

1.6 The Myth of the Given

Sellars's term "The Myth of the Given" has become something of a buzzword among contemporary philosophers who have taken themselves to have learned some lesson from Sellars, and it has become a cause of frustration for some contemporary philosophers who aren't part of the club that throws this term around but get it thrown at them. David Chalmers (2010), who gets this term thrown at him as much as anyone in contemporary philosophy, takes the term to pick out the view that "experiences have a special epistemic status that renders them 'given' to a subject," going on to say, "Sellars's (deliberately abusive) term for the view has caught on, and today it is not uncommon for this label to be used in criticizing such views as if no further argument is necessary," (299). I am quite sympathetic to Chalmers's frustration here. It is quite common for the label "the Myth of the Given" to be applied to some view in order to dismiss it without any further

[9] See Simonelli (2021) for a more thorough exegesis of Sellars according to which it is clear that worldly semantics falls within the scope of the Myth of the Given.

argument. Indeed, in many cases of its application, it seems to function as a *mere* label, without any clear descriptive content. The first thing I want to point out is that the set of views that are aptly characterized as instances of "the Myth of the Given" extends much more widely than specifically views about sensory experience. Sellars does start his critique of the Myth of the Given by considering views in which sensory experiences have a special epistemic status, but he takes this to be only "a first step in a general critique of the entire framework of givenness," (1956, 254). When I use the term, I mean to speak about this general framework.

There is considerable debate among commentators as to what, exactly, the general framework of givenness is. Some commentators, such as deVries and Triplett (2000), take the Myth to involve a rather particular version of foundationalism. On this way of construing things, Sellars's critique of the Myth would rather parochially pertain to views in epistemology that were prevalent in Sellars's day but are widely disregarded nowadays. As such, it'd be hard to see, on such a construal that the views in semantics that I am attacking here, which seem rather far removed from the epistemological foundationalism of the early part of the 20^{th} century, could be aptly characterized as instances of the Myth. Other commentators, such as Brandom (1997), construe the Myth as the view that there could be a form of non-conceptual awareness that directly entails having conceptual knowledge. Once again, on this way of construing the Myth, it pertains to a rather particular view in the philosophy of mind (and, indeed, one that few philosophers of mind nowadays would accept), and it's hard to see how it could pertain to the views in semantics that I am attacking here. So, if the Myth of the Given is not to be identified as deVries and Triplett or Brandom identify it, how is it to be identified? What *is* the Myth of the Given?

The answer to this question, I think, is surprisingly straightforward; the term "Myth of the Given" actually functions as perfectly sufficient as a description of what it picks out. The Myth of the Given is simply any conception of our knowledge of some aspect of reality as simply *given* to us, and intelligible only as given in this way. The basic problem with such a conception—what makes any such conception a *myth*—is that, by thinking of knowledge of some aspect of reality as given in this way, we preclude ourselves

from thinking of our knowledge as rational, and thus, as genuinely knowledge. Holding something rationally requires being able, at least in principle, to put it in to question and, in response to that question, articulate the reasons for holding it. If knowledge of some aspect of reality is taken to be simply given, and intelligible only as such, then this knowledge constitutes a stopping point in the inquiry into our knowledge of reality, at which no questions can be asked. But if no questions can be asked, then no reasons can be given, and so we cannot make sense of our knowledge of the aspect of reality that is supposedly given to us as rational. Accordingly, we cannot make sense of this supposedly given "knowledge" as genuinely knowledge. In other words, conceiving of knowledge of some aspect of reality as given to us, and intelligible only as such, undermines its very status as knowledge.

Stated in these terms, the problem with the Myth can seem obvious. However, many theories which attempt to explain our rational capacities can easily fall prey to it. The problem arises when a theory aims to explain our rational capacities as depending on knowledge that can really be understood only as an achievement of those very capacities. That is the basic problem with the sense data theory that Sellars addresses in the beginning of *Empiricism and the Philosophy of Mind*. The sense data theorist wants to conceive of our conceptual knowledge as asymmetrically dependent on our knowledge of sense data, but any knowledge of sense data that we might have can in fact be understood only as an achievement of our capacity for conceptual knowledge. This is a particularly clear case of the Myth, but, generally, one is led into the Myth when one attempts to give an account of some aspect of our rational capacities as dependent on some sort of knowledge or awareness that, given the explanatory project for which it is recruited, must be conceived of as not involving an actualization of the rational capacities that it does in fact involve. This problematic structure of Givenness is nicely stated by John McDowell (2009) as follows:

> Givenness in the sense of the Myth would be an availability for cognition to subjects whose getting what is supposedly Given to them does not draw on capacities required for the sort of cognition in question, (256).

If one is committed to such an explanatory structure, they face the following dilemma.

On the one hand, they can refuse to acknowledge the contribution of the capacities whose actualization is essentially involved in the knowledge to which they appeal, and thus are saddled with a view in which the sort of knowledge that they take to underlie the capacities they are trying to explain is *unintelligible*. On the other hand, they can acknowledge the capacities that are in fact essentially involved in the knowledge they appeal to in order to explain these very capacities, and thus are saddled with a view in which their account of these capacities is *incoherent*. This dilemma, which will arise repeatedly in various forms throughout this dissertation, is the basic way in which the Myth manifests itself in a dialogical context.

Now, McDowell, here and in much of his other work, follows Kant in emphasizing that our rational capacities are essentially *conceptual* capacities. Sellars is a Kantian that has undergone a linguistic turn. The crucial point for Sellars is that our conceptual capacities are essentially *linguistic* capacities. As he puts it "grasping a concept is always mastering the use of a word." Accordingly, if Sellars is right to follow Kant in thinking that awareness of anything of any cognitive significance requires the deployment of concepts, and the deployment of concepts is essentially a linguistic affair, then it follows that "all awareness of sorts, resemblances, facts, etc., in short, all awareness of abstract entities—indeed, all awareness even of particulars—is a linguistic affair" (1956, 289). This is Sellars's so-called "psychological nominalism."[10] With this claim on board, it is not hard to see what the problem for worldly semantics, from Sellarsian perspective, is. The problem for worldly semantics is that semantic competence is supposed to be explained as depending on knowledge of the world. This explanatory structure presupposed here requires that knowledge of the world *does not* presuppose semantic competence. However, insofar as this worldly knowledge is a product of rational capacities, rational capacities are essentially conceptual capacities, and grasping a concept is always mastering the use of a word, this knowledge *does* presuppose semantic competence. To put the problem in terms of McDowell's formulation of the problematic structure of Givenness, in articulating a worldly semantics for some language that some subjects speak, we require an

[10]Many have said that this is not a very good name for his position (Brandom, 1997). I think otherwise, but I won't get into such an argument here. For an explanation of the sense of Sellars's term and it's connection to the more familiar ontological sense of "nominalism," see Simonelli (2021).

availability for cognition of worldly knowledge to these subjects such that their getting this worldly knowledge does not draw on the capacities that are in fact required for having it, specifically, the capacity to use and understand sentences of that language.

Now, as I said above, to limit the scope of my discussion here, I have focused my attention on one aspect of our semantic competence that is supposed to be explained by a worldly semantic theory: our knowledge of facts consisting in sentences standing in certain relations of entailment and incompatibility in virtue of meaning what they do. On a worldly semantics, this semantic knowledge is taken to be asymmetrically dependent on worldly knowledge of what we might speak of the "metaphysical structure" reality. Worldly semantic frameworks, of both the extra- and intra-worldly variety are ultimately committed to a view in which the metaphysical structure of reality, be it articulated in terms of facts consisting in set-theoretic relations obtaining between extra-worldly entities or primitive modal relations obtaining between intra-worldly entities, is simply given to a potential learner of a language, such that that learner can map words, phrases, and sentences with their semantic values. Accordingly, they preclude us from being able to non-circularly comprehend the worldly knowledge which is supposed to underlie our knowledge of meaning as genuinely knowledge. Of course, so far, I have just stated this claim. I have not yet given any argument that worldly semantic theories are problematic in the way that I have claimed they are. That is what I will do in the next two chapters, arguing against the two most common incarnations of worldly semantics in the contemporary literature. Before I begin my attack on worldly semantics, however, I want to clarify my targets, especially "extra-worldly semantics," at which I will take aim in the next chapter.

1.7 Elucidatory and Explanatory Models in Semantics

Extra-worldly semantics is by far the most common type of semantics practiced by contemporary semantic theorists. In the next chapter, we'll see that, if taken to constitute an explanation of what it is in which speakers' knowledge of meaning consists, it involves a clear instance of the Myth. Many semantic theorists who employ such a framework, however, phrase what they are doing in such a way so as to not commit themselves to the

claim that they really are *explaining* the knowledge of meaning that speakers have. Rather, they often put things so as to suggest that they are doing something else: *elucidating* or *explicating* this knowledge of meaning. Rarely, however, are theorists explicit about what this distinction is or what falling on one side of it rather than the other amounts to.

Consider, for instance, what Gennaro Chierchia and Sally McConnell-Ginet (1990) say in their introductory semantics textbook when sensing possible trepidation from their scientifically-minded reader about the appeal to "possible worlds" in the semantic theory:

> [U]sing the formal framework of possible worlds in semantics has produced some very interesting and nontrivial accounts of various intensional phenomena, and many quite enlightening semantic studies have been generated. It certainly seems to us a fruitful hypothesis that our semantic competence can be elucidated in this framework, (207-208).

Chierchia and McConnell-Ginet say here that it seems to them to be a fruitful hypothesis that our semantic competence can be *elucidated* in the framework of possible worlds. They do not say that it seems to them that our semantic competence can be *explained* with the use of such a framework. Tellingly, when they originally lay our what their aim, as linguists, is, at the beginning of the book, they do so in such a way as to suggest that they are uncomfortable about the use of the expression "explain" in this context:

> [A]s linguists, our focus is on modeling the cognitive systems whose operation in some sense "explains" linguistic phenomena, (2).

They never say in *what* sense the operation of the cognitive systems they seek to model "explains" the linguistic phenomena with which they are concerned, and they never say why they put the expression "explains" in scare-quotes here. As they proceed in the book, this issue gets lost entirely. They drop this guardedness about the use of the expression "explains," and freely talk about the semantic theory "explaining" and "accounting for" empirical phenomena such as "judgments of semantic relatedness," using these expressions more or less interchangeably (51). This wavering between an elucidatory and an explanatory conception of semantics makes it difficult to determine to what extent the theories put forward by Chierchia and McConnel-Ginet are targets of the attack on extra-worldly semantics put forth in the next chapter. I do not deny that our semantic

competence can be *elucidated* with the use of a framework that centrally employs the notion of possible worlds. Indeed, I think it can be, and I think that this elucidation can indeed be enlightening. What I am denying is that our semantic competence can be *explained* or *accounted for* with the use of such a framework. This crucial distinction, I believe, is often lost in contemporary theorizing about meaning, and it is this tendency of contemporary theorists to lose this distinction that is largely responsible for the pervasiveness of the Myth in contemporary theorizing about meaning.[11]

One way to get the distinction between semantic theories that aim at elucidation and those that aim at explanation into view is to consider whether a certain sort of circularity is acceptable in the theory. A circular explanation, in which the facts that are supposed to be explained are appealed to in order to arrive at the "explanation" of them, is no explanation at all. If a semantic theorist is able to rest happily while being aware of circularity in their theory, then it is a good bet that they take themselves to be doing elucidatory rather than explanatory work. One such theorist is, Jaako Hintikka (1975), who, when considering the question of whether possible worlds could only be understood by reference to counterfactual claims which would then be understood in terms of a possible worlds semantics, writes,

> [A] circle of explication need not be a vicious one, provided it is wide enough to enable a logician to uncover nontrivial aspects of the structure of the concepts involved, (135).[12]

Here, Hintikka uses the term "explication" rather than the term "explanation," and this, of course, is no accident. A circle of *explanation* can only be a vicious one, but a circle of *explication* or *elucidation* need not be vicious. Stina Bäckström provides a clear statement of how at least some elucidations might be virtuously rather than viciously circular:

[11] This distinction between semantic elucidation and explanation is related to what is likely a more familiar distinction between *semantics* and *meta-semantics*, which is now standard in the literature (See Burgess and Sherman, 2014). I intentionally eschew this distinction in this dissertation; it does not mark a distinction that I am taking on board in my conceptualization of the issues here. It seems to me that this distinction is so systematically blurred in actual practice, where a formal semantic theory through which particular meanings are assigned almost always goes hand in hand with a proposal for what meaning in general is, that it yields more confusion than clarification.

[12] I should note, Hintikka himself does not "officially" make this claim. It actually occurs in the context of a fictional dialogue with Quine, but it is clear that it represents Hintikka's own view.

> There is [. . .] at least one kind of case in which a philosophical account can be circular without fault, and that is when the account aims at elucidating two concepts or phenomena that are mutually interdependent. In that case, circularity—far from being a deficiency—is a necessary feature of a successful account, (2016, 192).

If worldly knowledge and semantic knowledge are conceived of as mutually interdependent, then it is possible that they can be mutually elucidated by an account that appeals to one in explicating the other and vice versa. This is essentially how conceptions of possible worlds that define them with the use of "meaning postulates" (Carnap 1952, Partee 2005) suggest that we think of things (whether the proponents of such conceptions know that they are suggesting this or not). I will discuss such conceptions in more detail in the next chapter, but the point of bringing them up here is just to show what a self-consciously elucidatory rather than explanatory conception of worldly semantics could be. On such a conception, our semantic knowledge is explicated as depending on our worldly knowledge, but this worldly knowledge is analyzed in terms of our knowledge of relations among sets of possible worlds, and possible worlds are defined as depending on our semantic knowledge, made explicit by the laying down of "meaning postulates." Clearly, a possible worlds semantics that is structured in this way will not be able to explain or account for the knowledge of meaning that we have, since this knowledge must be appealed to in order to define the materials with which the theory is constructed, but it may very well be able to uncover non-trivial structural features of the meanings of which we have knowledge through the mutual elucidation of our semantic knowledge and our worldly knowledge.

If elucidating structural features of our semantic knowledge is all you are aiming to do in employing a possible worlds semantics, and you recognize the space for a semantic theory that actually explains this semantic knowledge, then you and I have no quarrel. There are many ends other than explanatory ones with which a possible worlds semantics that elucidates the structure of the space of meanings of which we have knowledge can aid us. Even in the context of an explanatory project, an elucidatory semantic theory can function to get the explanandum into view, and an explanatory theory can function to

explain it.[13] The great merit of an extra-worldly semantic framework is that it enables us to provide characterizations of semantic relations and semantic operations in set-theoretic terms. For instance, it enables us to provide a characterization of the semantic relations of entailment and incompatibility and semantic operations of conjunction and negation in terms of the set-theoretic relations of being a subset of and being disjoint and the set-theoretic operations of intersection and complementation. Once again, I don't deny that the structure of the semantic relations that complex sentences stand to one another can be elucidated set-theoretically by thinking of these sentences and their parts as having sets of possible worlds and related mathematical entities as their semantic values. What I deny is that to assign sets of possible worlds to sentences is to model their meaning in a way that enables us to explain that they stand in these semantic relations, our knowledge that they do, or our behavior that is a manifestation of this knowledge.

Because the distinction between elucidation and explanation gets lost, many semanticists take themselves to be explaining our knowledge of meaning when all they can really be doing is elucidating it, and this leads them to take there to be no room for semantic theories that really are of the sort to be able to explain it. For instance, Paul Portner (2005), in his introductory textbook to formal semantics, relevantly entitled *What Is Meaning?*, says that one of the main reasons for thinking of meaning in terms of a possible worlds semantics of the sort introduced there is that this sort of semantics "lets us define some basic semantic concepts: synonymy, contraiety, entailment, contradiction, tautology," (18). Portner makes this claim in the course of arguing that possible worlds semantics, as opposed to the holist empiricist semantics proposed by W.V. Quine (1953, 1960) or the social-normative semantics proposed by Robert Brandom (1994, 2000), is the sort of semantic theory with which we should think about what meaning is (Portner 2005, 4-22).[14] Possible worlds semantics, Portner claims on behalf of mainstream formal semantics, gives us a better account of what meaning is than the sort of semantic theories

[13] For elucidatory work in extra-worldly semantics that quite closely aligns with the Sellarsian explanatory project undertaken here, see Kraut (1979, 1982). I do not develop an account of the semantic phenomena explicated in those works, but, in providing a discursive role semantic analysis of them along Sellarsian lines, it's clear that Kraut's work can function as a helpful guide.

[14] It is worth noting that, while Portner drops a footnote to Quine's "Two Dogmas" and *Word and Object*, it is hard to see how the view he characterizes as Quine's really is Quine's. He seems to be failing to discriminate Quine's empiricist holism from Brandom's rationalist holism.

proposed by Quine or Brandom. If all possible worlds semantics is doing, however, is elucidating and not explaning, then possible worlds semantics does not give us a better account of what meaning is than the sorts of theories proposed by Quine or Brandom because it gives us no account of meaning at all.

By the end of this dissertation, I will have articulated a semantic theory that can actually account for what meaning is and what it is for us to grasp the meanings that we do. In contrast to the worldly semantic theories, on the theory I'll propose, knowledge of meaning is not undergirded by worldly knowledge. On the contrary, our "worldly" knowledge really is *nothing other than* our knowledge of meaning, expressed in a worldly mode. On this Sellarsian story, the "worldly knowledge" to which the worldly semanticist appeals, is conceived of as a "shadow" of our knowledge of meaning which is, in reality, not worldly, but normative. I will then provide a normative semantics where the behavioral patterns codified by these norms can be explained without appeal to knowledge of the worldly entities that the worldly semanticist takes to be contained in meanings. This will make space for an account of the real relation between language and the world, which, once again drawing from Sellars, I will provide. All of that in good time. But first things first: I must argue that worldly semantics, of both the extra- and intra-worldly variety, at least insofar as the semantic theories have explanatory rather than merely elucidatory aims, really do contain, at their very core, instances of the Myth of the Given.

2

Extra-Worldly Semantics

2.1 Introduction

In this chapter, I'll consider the explanatory potential of semantic theories that take the form of what I've called "extra-worldly semantics." In the paradigmatic case, such a semantic theory will be one in which we think of the meaning of a sentence in terms of the set of completely determinate ways for the world to be—the set of "possible worlds"—such that each element of that set is a way for the world to be such that that sentence is true.[1] As a variant of worldly semantics, an extra-worldly semantics requires us to try to comprehend our knowledge of meaning sentences and predicates as asymmetrically dependent on our knowledge worldly entities and their relations. Specifically, an extra-worldly semantics thinks of this knowledge as knowledge of possible worlds, objects contained within them, and set-theoretic relations between possible worlds and the objects contained within them. I will argue that we cannot comprehend our knowledge of meaning in this way. The core problem is that the extra-worldly knowledge to which an extra-worldly semantics appeals is only intelligible as *dependent on* our knowledge of propositions and properties, but this knowledge of propositions and properties, on an extra-worldly semantics, is *understood in terms of* our knowledge of sets of worlds and functions from worlds to extensions. I will go on to argue in the next chapter that this knowledge of propositions and properties is itself dependent on our knowledge of the correct use of sentences and predicates, but it suffices for my purposes in this chapter to show that knowledge of propositions and properties cannot be understood in terms of knowledge of possible worlds and the

[1] Or, equivalently as a function from the total set of worlds to the value *true* or *false*. Informational dynamic semantic theories, which I'll discuss briefly in Chapter Four, are also variants of extra-worldly semantics.

objects contained within them. In the characteristic form of the Myth of the Given, an extra-worldly semantics requires the availability for extra-worldly knowledge to subjects whose getting this extra-worldly knowledge does not draw on the capacities required for this exra-worldly knowledge.

2.2 The Extra-Worldly Meaning of Predicates

In the previous chapter (Section 1.2), I introduced a toy language consisting in sentences like "*a* is gray," "*b* is darker than *c*," "It's not the case that *c* is white," "*c* is gray or *b* is black," and so on. Our task is now to articulate a semantic theory for this language that is able to explain the behavior that speakers of it exhibit in virtue of grasping the meanings of the expressions that belong to it. More concretely, our task is to construct a function $[\![\cdot]\!]$ that maps each sentence φ of this toy language to its semantic value $[\![\varphi]\!]$, the element of our theory that is meant to serve as a model of what speakers of this language know in knowing the meaning of φ and to do so in such a way that the semantic value of a complex sentence is determined by the semantic values of its parts and the way those parts are put together. The formal framework in which this task is usually undertaken is what I am calling an "extra-worldly" semantic framework, or what is more commonly called a "possible worlds semantics," and that is the formal framework that we'll consider in this chapter.

There are several motivations for an extra-worldly semantics. Perhaps the main motivation is that it functions as a single framework in which we can provide set-theoretic characterizations of various phenomena of concern to the study of linguistic meaning: semantic relations such as synonymy, entailment, and incompatibility, logical operations such as conjunction, disjunction, and negation, modal notions such as possibility, necessity, permission, and obligation (Kripke 1959, 1980; Kratzer 1977), epistemic notions such as knowledge and belief (Hintikka 1962), counterfactuals (Lewis 1973), pragmatic phenomena such as the effect of making an assertion on a state of inquiry (Stalnaker 1978).[2] All of this, however, hangs on the thought that simple content words such as the basic

[2]For an overview of the uses of possible worlds in semantics, see Partee (1988).

predicates of a language can be reasonably assigned semantic values in a possible worlds semantics. It is the assignment of semantic values to simple content words that principally concerns us here, so it is worth considering how a possible worlds semantics can be and often is motivated by the claim that such a semantic theory enables us to provide adequate assignments of semantic values to simple content words.

Consider first an *extensional* semantic theory of the sort proposed in Heim and Kratzer's (1998) introductory semantics textbook. On such a theory, we take the semantic values of names to be objects and the semantic values of predicates to be the sets of objects to which those predicates apply.[3] It does not take long to see that extensional semantic values cannot possibly serve as adequate models of what speakers grasp in grasping the meanings of predicates. To see this, consider the following example. Suppose just three individuals smoke, Joe, Mary, and Sue. In which case, following two equations both specify the semantic value of "smokes":

$[\![\textbf{smokes}]\!] = \{x : x \text{ smokes}\}$
$[\![\textbf{smokes}]\!] = \{\text{Joe, Mary, Sue}\}$

Now, suppose it just so happens that Joe, Mary, and Sue are also the only individuals who ski. In which case, the following two equations both specifiy the semantic value of "skis"

$[\![\textbf{skis}]\!] = \{x : x \text{ skis}\}$
$[\![\textbf{skis}]\!] = \{\text{Joe, Mary, Sue}\}$

As you can see, in this hypothetical scenario, the semantic value of "smokes" is identical to the semantic value of "skis." It is just the set of these three individuals. Clearly, however, in such a scenario, "smokes" would not mean the same thing as "skis;" "smokes" would still mean *smokes* and "skis" would still mean *skis*. It seems that our semantic theory ought not have the consequence that, if it just so happens that the same people who smoke are the ones who ski, the words "smokes" and "skis" would mean the same thing. Heim and Kratzer, of course, don't want their theory to have this consequence. In response to this problem, they say that, even though the expressions on the right of each of the two

[3]Or the characteristic functions of such sets, as Heim and Kratzer officially formulate things. These formal details don't matter to the conceptual point here.

equations that are candidate specifications of the semantic values of "smokes" and "skis" define the same set, only the first type of expression is the sort that should go into our semantic theory. Only if we state the set that is the extension of the predicate with the use of a *condition*—specifying the set that is the semantic value of "smokes" or "skis" as the set of things that *smoke* or the set of things that *ski*—do we state that expression's semantic value in a way that "shows" its meaning. If we specify the set that is the extension of a predicate by merely listing its members, we do not show its meaning. The upshot of this response is to deny that the meaning of a predicate is to be identified with its semantic value. It is not sufficient, in the context of a semantic theory, to specify the semantic value of a predicate; one must specify it *in a particular way*, a way that "shows" its meaning.

In his discussion of the theoretical role that semantic values are to play, Yalcin (2018) takes issue with Heim and Kratzer's response to this problem and proposes an alternative:

> The conclusion to draw from the problem they raise is not that meaning must reside somewhere beyond semantic value; it is that the semantic values initially postulated are not fruitful, because too coarsegrained. We need richer semantic values to capture the sorts of distinctions that need distinguishing [. . .] A better response to the problem would simply be to introduce intensional resources (possible worlds or situations) at the start, as a beginning at fixing the problem, at delivering a semantic theory that can make at least a minimal range of the distinctions between semantic values that need to be distinguished, (339).

Insofar as semantic values are the entities in theory that are meant to model the meanings of which speakers have knowledge, a theory that assigns the same semantic value to two predicates that differ in meaning but happen to have the same extension is inadequate. This inadequacy can be rectified, Yalcin suggests, by appealing to possible worlds from the outset in assigning semantic values to predicates. This alternative option is undertaken in several other introductory semantics textbooks, for instance, Kate Kearns' (2011) textbook. There, Kearns originally introduces the notion of possible worlds by considering the inadequacy of taking the extensions of predicates—the sets of objects to which they apply—to be their semantic values. She considers a theory that takes the semantic value of "brown" to be the set of brown things in the world, and she writes,

> Is that all there is to it? Suppose the world was exactly the way it is except for one detail—a certain brown pottery bowl on a windowsill in Ladakh is blue

> instead of brown. If the world was like that instead of how it is, then the set of brown things would be different, but surely the word *brown* wouldn't have a different meaning. This seems to make the word meaning depend on accidents of fate.
>
> We want to take into account the way the word *brown* would relate to the world even if things were a bit different from the way they actually are. We want to take into account not only the objects a predicate happens to apply to in fact, but also all the hypothetical objects that it would apply to, meaning what it does mean, if things were different. [. . .] We need to consider hypothetical versions of the whole of reality to state what individual predicates would apply to in virtue of their meaning. Words connect not only with the real world, but also with other possible worlds, (7).

So, rather than taking the meaning of a predicate to be its *extension*, we can take the meaning of a predicate to be an *intension*, a function from possible worlds to extensions. The meaning of "brown" will thus be a function that maps each possible world to the set of things that are brown in that world. So, the meaning of "brown," in the possible world Kearns describes, in which the pottery bowl of which she speaks is blue, is not different than it is in the actual world in which it is brown. In both possible worlds, "brown" still semantically expresses the function that maps each possible world to the set of brown things in that world. Though the set of brown things varies between the actual world and the possible world that Kearns describes, the function expressed by "brown" is invariant across these two possible worlds. We thus avoid the problem with the extensional theory of Heim and Kratzer, without having to say that meaning resides somewhere beyond semantic value.

The basic thought underlying this way of thinking about the meanings of predicates is that to grasp the meaning of a predicate is to grasp a rule for sorting things into the things to which the predicate applies and the things to which it does not. More determinately, to grasp the meaning of a predicate is to grasp a rule for that enables one to take any possible way for things to be, any possible world, and sort things, as they are in that world, into two sets: the set of things to which the predicate applies and the set of things to which it does not. Accordingly, the meaning of a predicate can be modeled as a function that maps each possible world to the set of things that satisfy the predicate in that world. Thus, for instance, the meaning of "brown" can be modeled as a function that maps each world to

the set of things that are brown in that world. Such functions are, at least in the context of such an extra-worldly framework, thought of as the properties that are expressed by those predicates. For instance, Portner (2005) says, "a property can be thought of as an association between worlds and sets—it provides, for each world, a set of things," (55-56). So, the property of being brown is thought of as being the function that is the semantic value of "brown." What speakers grasp in grasping that "brown" and "pink" are incompatible is that, for any possible world, the set of things that are brown and the set of things that are pink are disjoint. No matter how things are, if something is pink, then it isn't brown, and vice versa.

Extra-worldly semantic theorists are often not particularly explicit as to whether properties, those entities that are semantically expressed by 1-place predicates, *really are* functions from possible worlds to extensions, or whether properties are simply *adequately modeled by* functions from possible worlds to extensions, and likewise for the analogous question with respect to propositions and whether they are to be identified with sets of possible worlds (or functions from worlds to truth values). When theorists are explicit, their answers tend to vary. Andy Egan (2004), for instance, proposes *identifying* properties with functions from worlds to extensions. Other theorists are explicit that they do not wish to make this identity claim. Commenting on the analogous question for propositions, Chierchia and McConnell-Ginet write

> We are not claiming that sets of worlds are what propositions really are. We claim only that sets of worlds have got the right structure to do what we think propositions do: mediate between sentences and their truth conditions. A crucial component of our semantic competence, what we understand when we understand the content of a sentence, is our capacity to match sentences and situations. Functions from worlds to truth values can be regarded as an abstract way of characterizing such a capacity, (211).

When I get to my main argument in Section 2.6, I will argue that, in order for our knowledge of the meanings of predicates to be *adequately modeled* by functions from worlds to extensions, what we actually know in knowing the meaning of a predicate must *really be* something quite close to such a function. It's important to be clear from the outset, however, that this is a claim that I'm making in the course of arguing against the

weaker claim that functions from worlds to sets of objects cannot really be *adequate models* of what we grasp in grasping the meaning of a predicates. To put this all in perspective, on the semantic theory I will eventually defend, the semantic value of a sentence will be a function that maps each discursive context in which someone might employ that sentence to the discursive context that would result upon their employing it. I do not claim that the meanings of sentences *really are* such functions, but I do claim that the meanings of sentences can be *adequately modeled* by such functions. That is what I am claiming is not so of the semantic values provided by the extra-worldly semantic framework.

2.3 A Simple Extra-Worldly Semantics

To investigate the structure of an extra-worldly semantic framework more carefully and systematically, let us turn again to our toy language and lay out a simple extra-worldly semantic theory for it. We start with a model $\langle W, U, V \rangle$ consisting in a set of worlds W, a set of objects U, and a valuation function V. Let us consider first the first two elements of the model. In an extra-worldly semantic framework, we start with the notion of a completely determinate way for the world to be: a "possible world." The way the world actually is, of course, is one completely determinate way for the world to be. But there are different ways that the world could have been that are other than the way that it actually is. So, there are, we might say, other "possible worlds," worlds that are not actual, but merely possible. Applying this idea to the world in which our toy language is spoken and assuming, for the purpose of simplicity, that there can be only the three objects contained in it, there are twenty-seven completely determinate ways for the world to be—twenty-seven possible worlds. There is, for instance, the world in which a is gray, b is white, and c is black, there is the world in which a is gray, b is gray, and c is white, and so on. W is the set of these twenty-seven possible worlds, and U is just the set of the three objects contained in each of these possible worlds, a, b, and c.

Now, consider the valuation function V. This function assigns to a name a function that maps each possible world to a particular object (the one that is actually named by that name), assigns to a 1-place predicate a function that maps each possible world to a

set of objects (the ones that satisfy the predicate in that world), and assigns to a 2-place predicate a function that maps each possible world to a set of pairs of objects (the pairs that satisfy the predicate in that world). So, our valuation function will assign to the name "*a*" a function that maps each possible world to *a*, it will assign to the predicate "is gray" a function that maps each possible world to the set of things that are gray in that world, it will assign to the predicate "is darker than" a function that maps each possible word to the set of pairs of objects such that the first is darker than the second in that world, and so on. It thus assigns meanings to the basic content words of the language for which we are constructing a semantic theory; it tells us what objects are named by the names of our toy language, which properties are expressed by the 1-place predicates of our toy language, and which relations are expressed by 2-place predicates of our toy language. This constitutes the ground level of the semantic theory.

Having specified a model of this sort, our aim in constructing a semantic theory is to devise a way of assigning semantic values to all of the complex expressions of our language on the basis of the valuations that our model assigns to simple ones. So, for name n, and for a 1- or 2-place predicate P, the value is just what our model gives us:

1. $[\![n]\!] = V(n)$
2. $[\![P]\!] = V(P)$

Once we've specified these values, we can assign a value to any sentence consisting in a name followed by a 1-place predicate, and a name followed by a 2-place predicate and then another name as follows:[4]

1. $[\![nP]\!] = \{w : [\![n]\!](w) \in [\![P]\!](w)\}$
2. $[\![n_1Pn_2]\!] = \{w : ([\![n_1]\!](w), [\![n_2]\!](w)) \in [\![P]\!](w)\}$

So, some world w is an element of $[\![nP]\!]$ just in case the object to which $[\![n]\!]$ maps w is an element of the set of objects to which $[\![P]\!]$ maps w. And some world w is an element of $[\![n_1Pn_2]\!]$ just in case the pair of objects consisting in the object to which $[\![n_1]\!]$ maps w and

[4]This notation—specifically, using functional notation of the form $f(x)$ with semantic values as the functions and worlds as arguments—is non-standard. I have put things this way to make the basic structure of the theory particularly clear.

then the object to which $[\![n_2]\!]$ maps w is an element of the set of pairs of objects to which $[\![P]\!]$ maps w. The result of these composition rules is that atomic sentences of our toy language are assigned some subset of these possible worlds as semantic values. Consider, for instance, the set of worlds that will be assigned to "a is gray." The semantic value of a is a function that maps each possible world to a, and the semantic value of "is gray" is a function that maps each possible world to the set of gray things in that world. The semantic value of "a is gray," then, will be the set of worlds such that a is an element of the set of gray things in those worlds. That is, it will be the set of worlds in which a is gray. Likewise, the semantic value of the sentence "a is darker than b" will be the set of worlds in which a is darker than b, the semantic value of the sentence "a is the same color as a" is the set of all twenty-seven possible worlds, since every world is such that each object in it is the same color as itself, the semantic value of the sentence "a is lighter than a" is the set of no possible worlds, and so on.

Once we've assigned values for all the atomic sentences in this way, we can assign values to logically complex sentences with the use of set-theoretic operations of complementation, intersection, and union as follows:

2. $[\![(\textbf{It is not the case that } \varphi)]\!] = W - [\![\varphi]\!]$
3. $[\![(\varphi \textbf{ and } \psi)]\!] = [\![\varphi]\!] \cap [\![\psi]\!]$
4. $[\![(\varphi \textbf{ or } \psi)]\!] = [\![\varphi]\!] \cup [\![\psi]\!]$

Assigning values to complex sentences in this way, we have a semantics with which we can assign semantic values to all of the sentences of our toy language, all infinity of them. So, for instance, the set of worlds assigned to "It's not the case that c is gray," will be the set of all the worlds that are not elements of the set of worlds in which c is gray. The set of worlds assigned to "a is lighter than b or b is white" will be the set of worlds that are either an element of the set of worlds in which a is lighter than b or an element of the set of worlds in which b is white. The set of worlds assigned to "(It's not the case that c is gray) and (a is lighter than b or b is white)" will be the intersection of the first set and the second set. And so on.

We now have a function $[\![\cdot]\!]$, defined for all of the sentences of our toy language, such that, for any sentence of our toy language φ, $[\![\varphi]\!]$ is a formal model of the meaning of φ. We have a simple semantic theory for our toy language. Now that we've constructed a simple semantic theory, let's see what theoretical work it can do. Consider again the following set of facts:

F1. The sentence "a is darker than b" is synonymous with the sentence "b is lighter than a."

F2. The sentences "a is black" and "b is gray" jointly entail the sentence "a is darker than b."

F3. The sentence "a is gray" is incompatible with the sentence "a is white."

As we've said, these are the sort of facts for which we want our semantic theory to account. Our guiding idea in constructing a semantic theory that can account for these facts is that speakers of a language behave in certain ways because they know that certain sentences of their language are synonymous with one another, entail one another, or are incompatible with one another, and they have this knowledge because they know what these sentences mean. If, by assigning meanings to these sentences, our semantic theory enables us to account for facts like (F1)-(F3), then, by taking speakers to have knowledge of meanings we assign to these sentences, we can explain their knowledge of these facts, thereby explaining their behavior as a manifestation of this semantic knowledge. The simple possible worlds semantics we've just sketched promises to enable us to do this. Let's see how.

On the simple semantic theory just sketched, we can give the following definitions. First, two sentences, φ and ψ, are synonymous just in case the set of worlds that is the value of φ is identical to the set of worlds that is the value of ψ. That is, φ is synonymous with ψ just in case $[\![\varphi]\!] = [\![\psi]\!]$. So, "$a$ is darker than b" is synonymous with the sentence "b is lighter than a" just in case every world that is an element of the set of worlds assigned to "a is darker than b" is also an element of the set of worlds assigned to "b is lighter than a," and vice versa. Since that is indeed so, (F1) obtains. Second, two sentences, φ and ψ, jointly entail another sentence, χ, just in case the intersection of the sets of worlds that is the value of φ and the set of worlds that is the value of ψ is a subset of the set of worlds

that is the value of χ. That is, two sentences φ and ψ jointly entail another sentence, χ just in case $([\![\varphi]\!] \cap [\![\psi]\!]) \subseteq [\![\chi]\!]$. So, "$a$ is black" and "b is gray" jointly entail "a is darker than b" just in case any world that is an element of both the set of worlds assigned to "a is black" and the set of worlds assigned to"b is gray" is an element of the set of worlds assigned to "a is darker than b." Since this is so, (F2) obtains. Finally, two sentences φ and ψ are incompatible just in case the sets of worlds that are their values are disjoint. That is, φ is incompatible with ψ just in case $[\![\varphi]\!] \cap [\![\psi]\!] = \varnothing$. So, "$a$ is gray" is incompatible with "a is white" just in case there is no world that is an element of both the set of worlds assigned to "a is gray" and the set of worlds assigned to "a is white." Since there is no such world, (F3) obtains. With these definitions, it seems that the simple possible worlds semantics just sketched enables us to account for (F1), (F2), and (F3).

Given that we can account for (F1)-(F3), it seems that, by modeling speakers' knowledge of the meaning of the sentences "a is gray," "b is white," "a" is white," "a is darker than b," and "b is lighter than a" as knowledge of the semantic values that our semantic theory assigns to them, we can explain speakers' knowledge of these facts, thereby explaining the behavior they exhibit in virtue of having this knowledge. Consider, for instance, the fact that competent speakers of our toy language behave in a way that manifests their knowledge of the fact that the sentences "a is gray" and "a is white" are incompatible. Recall, they never utter both sentences at the same time, they correct incompetent speakers that do, and so on. The explanation of this behavior, on this model, is that they know that the sets of worlds that are the values of these two sentences are disjoint. Uttering "a is gray" would function to inform other speakers that the actual world is among the set of worlds in which a is gray. Uttering "a is white" would function to inform other speakers that the actual world is among the set of worlds in which a is white. The knowledge of the incompatibility of these two sentences that competent speakers have consists in their knowledge that the sets of worlds that are the values of these two sentences are disjoint. Having this knowledge, they know that uttering both sentences would function rule out every possible world. Knowing this, they know to never utter both sentences at the same time, to correct incompetent speakers that do, and so on. In this way, it seems that our simple extra-worldly semantics enables us to explain the behavior we set out to explain.

Things, however, are not how they seem. To see why this is so, let us turn to the core notion of a possible worlds semantics: the possible world.

2.4 The Issue of Defining Possible Worlds

In a possible worlds semantics, a possible world w is often officially defined as a function that maps each sentence in the set of atomic sentences $\mathcal{A}$ to one of two values, *true* or *false* (Dever 2012, 51). There are other formally interchangeable definitions, but I'll call this one the "standard definition."[5] Officially, the standard definition is the following:

> A possible world w is any function $f : \mathcal{A} \rightarrow \{true, false\}$.

The intuition behind this definition is clear enough. A possible world is something that determines, for each atomic sentence of the language, whether that sentence is true or false. Accordingly, a possible world w can be defined as a function that maps each atomic sentence to a value, *true* or *false*. Having defined possible worlds as these functions, we can officially say what it is for an atomic sentence to be true in a world as follows:

> For any atomic sentence p, p is true in w just in case $w(p) = true$

This enables us to officially assign truth values to atomic sentences relative to possible worlds at the base level of our semantic theory, and then we can go from there.

Though the standard definition is widely treated as good enough for the purposes of laying down the groundwork for a possible worlds semantics, it does not take long to see what is wrong with it. Not only does it give us possible worlds, but it gives us "worlds" that are not possible as well. For instance, "a is gray" and "a is white" are both atomic sentences, and so there is a function $f : \mathcal{A} \rightarrow \{true, false\}$ that maps "a is gray" to *true* and maps "a is white" to *true*. On the standard definition, this gives us a "possible world," one in which "a is gray" is true and "a is white" is true. But clearly there is no possible world in which "a is gray" is true and "a is white" is true. If a is gray, then it can't be the case that a is white. So, there is no possible world in which "a is gray" is true and "a is

[5] We could equally define a possible world as a subset of the set of atomic sentences $\mathcal{A}$, and then the standard definition would be the characteristic function of that set (Veltmann 1996, 228).

white" is true. But the standard definition says there is. That's a problem. This problem, though completely obvious, turns out to be critical.

A first response is to offer a revised definition. One might, for instance, start by specifying which sets of atomic sentences are incompatible and then say that a possible world is a function that maps each atomic sentence of the language to a value *true* or *false* in such a way that it does not map all the members of any such set to the value *true*. This excludes a function that maps both "*a* is gray" and "*a* is white" to *true* from being a possible world, since the set consisting of "*a* is gray" and "*a* is white" is a set of incompatible sentences. The problem with saying this, however, is that our simple semantic theory was supposed to enable us to *account for* the fact that the sentence "*a* is gray" is incompatible with the sentence "*a* is white." Defining possible worlds in such a way that they depend on this fact precludes us from being able to do this. As we've said, on the simple semantics we've given, two sentences φ and ψ are incompatible just in case $[\![\varphi]\!] \cap [\![\psi]\!] = \varnothing$. So, "*a* is gray" is incompatible with "*a* is white" just in case there is no world that is an element of both the set of worlds in which "*a* is gray" is true and the set of worlds in which "*a* is white" is true. Is there any such world? Well, on the standard definition there is; on the revised definition, there is not. However, the reason *why* there is no such world on the revised definition is that "*a* is gray" is incompatible with "*a* is white," so the revised definition does not count any function that maps both sentences to the value *true* as a world. Since the fact that "*a* is gray" is incompatible with "*a* is white" *explains why* there is no world that is an element of both the set of worlds in which "*a* is gray" is true and the set of worlds in which "*a* is white" is true, saying that the two sets are disjoint cannot amount to giving an account of *what it is* for "*a* is gray" to be incompatible with "*a* is white." Schematically, if the fact that *A explains* the fact that *B*, then the fact that *A* cannot *just consist in* the fact that *B*. This is the first instance of a principle to which we will return several times throughout this dissertation. Here, the upshot should be quite clear: If we adopt the revised definition, the "account" of incompatibility given by possible worlds semantics cannot be an account at all.

A more sophisticated response to our problem, owed to Rudolph Carnap (1952) and advocated in contemporary semantics by Barbara Partee (2005), is to say that, in order

to properly define the space of possible worlds, we must lay down certain "meaning postulates" which function to constrain the model on which we base our semantic theory so that it includes only genuinely possible worlds. We said that our valuation function, applied to a 1-place predicate, gives us a function that maps each possible world to the set of objects that satisfy that predicate in that world. What we need to do is restrict the set of possible worlds that we let into our model by specifying which predicates can't be jointly satisfied by a single object in a world, which predicates are such that, if they are satisfied by some objects in some world, require other predicates to be satisfied by those objects in that world, and so on. Here, according Carnap and Partee, is where "meaning postulates" come in. The idea is that if we want to constrain which "worlds" get included in the model on which we base our possible worlds semantics, we can do this by laying down something like the following:

$$\forall x(\textbf{gray}(x) \rightarrow \neg \textbf{white}(x))$$

Here, **gray** is the symbol that we're using in our semantic theory to symbolize the predicate "is gray" of our toy language, and **white** is the symbol that we're using for the predicate "is white." This postulate says that, for any object x, if x satisfies the predicate "is gray," then it is not the case that x satisfies the predicate "is white." Laying down this postulate enables us to put a constraint on which "worlds" get counted as worlds in our model—it enables us to rule out any "world" w in which there is some object x, such that x is an element of $V(\textbf{gray})(w)$ and an element of $V(\textbf{white})(w)$. We are thus able to formally capture the fact that the predicates "is gray" and "is white" are incompatible in our semantic theory by making it such that the model on which it is based contains no world w in which there is an object x that satisfies both predicates.

It is now crystal clear, however, that our possible worlds semantics, based on a model that is determined by meaning postulates of the above sort, does not and cannot give us an account of either the fact that the predicate "is gray" and the predicate "is white" are incompatible or of our knowledge of that fact. Our semantic theory only "captures" this fact because we laid down the meaning postulate that we did, and we laid down this meaning postulate only because we know that the predicate "is gray" and the predicate "is

white" are incompatible. So, since our knowledge of the fact that "is gray" is incompatible with the predicate "is white" *explains why* our semantic theory contains no world in which there is some object x that satisfies both predicates, we cannot *account for* this fact or our knowledge thereof with the use of this semantic theory. This is essentially the same issue as with the revised definition; it just arises here at the level of predicates rather than at the level of sentences. In order to define possible worlds, we must appeal to our knowledge of the very facts that our possible worlds semantics was meant to explain. Of course, if our aim is just to elucidate our semantic knowledge rather than to account for it, then there is no problem here. As discussed earlier, however, contemporary semantics, at least as it is often advertised, has more than merely elucidatory ambitions.

It's worth noting that Carnap himself, the godfather of extra-worldly semantics, is quite clear about the fact that he is not aiming to account for speakers' knowledge of meaning. Commenting on what grounds the theorist's writing down certain meaning postulates, he writes,

> How does [the theorist] know that these properties are incompatible and that therefore he has to lay down postulate P_1? This is not a matter of knowledge but of decision. His knowledge or belief that the English words 'bachelor' and 'married' [or 'gray' and 'white' in our case] are always or usually understood in such a way that they are incompatible may influence his decision if he has the intention to reflect in his system some of the meaning relations of English words, (1952, 68).

Here, Carnap says that if we want our system to *reflect* the fact that certain words "are always or usually understood in such a way that they are incompatible," we'll lay down certain meaning postulates rather than others. He is under no illusion that a semantic theory of the sort he is proposing will be able to *account for* the understanding of the incompatibility of certain English words that English speakers have; it will simply reflect this understanding. In other words, Carnap's ambitions here are self-consciously *elucidatory* rather than *explanatory*. In *Meaning and Necessity*, he's clear that his main aim in providing the semantic analyses that he does is the *clarification* of philosophical concepts, aiming to replace the vague concept of, say, a sentence's being necessarily truth, with the precise concept of a sentence's holding in every state description (1947, 7-13). Carnap's aims,

however, are not those of contemporary linguistic theorists, who often do take themselves to be engaging in a genuinely explanatory enterprise.

Insofar as our aims are more than merely elucidatory, possible worlds cannot be defined in terms of sentences or predicates of the language for which one is constructing a semantic theory, for doing so requires one to the very semantic knowledge that is supposed to be accounted for by the theory. I take it that, by and large, those who employ talk of possible worlds with genuinely explanatory ambitions will not resist this conclusion. By such a theorist's lights, a possible world is simply a completely determinate way for the world to be. The set of possible worlds that there is does not depend on the semantic relations that obtain between expressions of a language, but, rather, simply on the set of possible worlds that there really are or on the set of ways that the world can possibly be. Knowledge of possible worlds thus does not require semantic knowledge of the sort that a possible worlds semantics seeks to explain. Rather, it is simply knowledge of the set of possible worlds that there really are or of the set of ways that the world can possibly be. So, the knowledge that underlies the knowledge of the fact that the sentence "*a* is gray" is incompatible with sentence "*a* is white" is either the knowledge that there is no world in which *a* is gray and *a* is white or that the world cannot be such that *a* is gray and *a* is white. That's not a fact about meaning but a fact about the world or, perhaps, the worlds.

Now, at this point, once one takes knowledge of possible worlds to be knowledge of non-linguistic worldly entities rather than knowledge of linguistic entities and their semantic relations, there are two ways to go: one can take possible worlds to be *composed* out of other worldly entities, such as propositions, states of affairs, or properties, as Adams (1974) and Plantinga (1976) classically do and Soames (2010) and King (2007a) more recently do, or one can take possible worlds to be *primitive* worldly entities. I will put off discussion of the former proposal for the next chapter, which concerns intra-worldly knowledge, and consider here the primitivist proposal, which takes knowledge of possible worlds to be a basic sort of worldly knowledge, not derivative on intra-worldly knowledge. An extra-worldly semantics proper, the sort of semantic theory proposed by Lewis and Stalnaker, takes our semantic knowledge to be based on worldly knowledge that is properly extra-worldly.

2.5 The Primitivist Proposal

The definitions of possible worlds that we've just considered aim to define possible worlds in terms of expressions of the language for which we're giving a possible worlds semantics. David Lewis, one of the philosophical pioneers of extra-worldly semantics, rejects this sort of approach, and he does so largely because of the issue with which we're concerned. He writes,

> [I]t would do us nothing to identify possible worlds with sets of sentences (or the like), since we would need the notion of possibility otherwise understood to specify correctly which sets of sentences were to be identified with worlds, (1973, 86).

Lewis recognizes here, that, if we identified possible worlds with formal constructions from sentences, we could not then use a semantics based on possible worlds in order to give an account of the notion of possibility. In order to say which sets of sentences are to be identified with worlds, we'd need to say which sets of sentences are compossible, and to do that would be to appeal to the very modal notions that we're trying to account for with the use of possible worlds. Now, Lewis's principal concern is with giving a semantics for modal sentences, but the same issue applies in our attempt to account for the modally robust semantic relations that obtain between ordinary non-modal sentences. Though our example has led us to focus on the notion of *in*compossibility here—the incompatibility of the sentences "*a* is gray" and "*a* is white"—the issue is just the same. It is the issue that precludes us from being able to use any of the accounts of possible worlds considered thus far if we're going to attempt to employ a possible worlds semantics to give an account of what it is for two sentences to be compossible or incompossible—compatible or incompatible. That's our issue. Lewis proposes a novel solution to it. Rather than thinking of possible worlds as sets of sentences or functions from sentences to truth values, Lewis opts to think of them directly by analogy to the actual world (Lewis, 1973, 1986).

Lewis's approach is a "primitivist" one: we do not try to say what a possible world is in terms of entities that are not possible worlds. By Lewis's lights, we know what sort

of thing the actual world is, and that's a possible world, so it's sufficient to say that other possible worlds are other entities just like the actual world. To say that this world is the "actual" one, on his view, is not to claim that there is a special property of existence or reality that only this world has. For Lewis, "actual" doesn't mean existent or real. Rather, he thinks of it as an indexical like "here" or "now." Just like "now," when uttered by some speaker at some time, just picks out the time at which one happens to be speaking, "actual," when uttered by some speaker in some world, just picks out the world in which one happens to be speaking. Just as thinking of "now" as an indexical opens the door for the view that times other than the one we happen to be in are equally real, thinking of "actual" as an indexical opens up the door for genuine realism about possible worlds, the view that possible worlds other than the one that we happen to be in are equally real. Having opened this door, Lewis, in what can be described only as an act of intellectual bravery, walks through it.

So, Lewis's way of being a primitivist about possible worlds is to be a genuine realist about them. Most theorists, however, have not wanted to walk through this door with Lewis. Perhaps the most prominent such theorist is Robert Stalnaker, another philosophical pioneer of extra-worldly semantics. Stalnaker (1986), endorsing what he calls "modest realism," maintains with Lewis that possible worlds aren't to be defined in terms of things other than possible worlds, but he does not go all the way to the genuine realism of Lewis. The view starts with the thought that there are two ways in which one might take there to be "many ways things could have been besides the way they actually are," (Lewis 1973, 84). On the one hand, one might take this world to be what we're speaking of when we speak of "the way that things actually are," and take there to be other worlds, other entities just like this one, that we are speaking of when we speak of "other ways that things could have been." On the other hand, one might think that what we're speaking of when we talk of "the way things actually are" isn't the actual world *itself*, but *the way the actual world is*: the property that the actual world instantiates in being just the way that it actually is. This, Stalnaker thinks, is an important distinction. Making it, we are able to maintain that there are ways that things could have been, without thinking that they are the same sort of thing that the actual world is. Possible worlds aren't concrete objects, but

abstract objects: properties, the sort of thing that objects instantiate, rather than objects themselves (objects, that is, which are not properties). While the actual world itself is not a property, the way the actual world is *is* a property, the property that the actual world instantiates in being just the way that it is. If the actual world were some other way, then, being this other way, it would instantiate some other property, some other way for the world to be. On Stalnaker's view, there actually are all of these other properties; they are "possible worlds" (or, perhaps less misleadingly called, "possible world-states") that figure into our semantic theory.

Before I go on to argue against the application of these primitivist conceptions of possible worlds in our semantic theory, I want to briefly consider one property that Stalnaker thinks his modestly construed possible worlds might reasonably be taken to have that makes them better candidates than Lewis's genuine other worlds. Stalnaker takes it that his worlds can reasonably be taken to be such that their existence depends on the activities of language speakers. On the basis of saying that possible worlds are "abstract objects whose existence is inferred or abstracted from the activities of rational agents," Stalnaker says "It is thus not implausible to suppose that their existence is in some sense dependent on, and that their natures must be explained in terms of, those activities," (1984, 51-52). This suggestion of Stalnaker's is cited approvingly by many theorists of meaning who wish to employ the framework of possible worlds in order to give an account of the meanings of the expressions of a natural language without having to commit themselves to a potentially scientifically questionable sort of realism about these entities (Chierchia and McConnell-Ginet 1990, 207-208; Partee 1988, 102). I take it, however, that this suggestion of Stalnaker's, in conjunction with the explanatory aim of semantics, is incoherent. One place where this incoherence arises is in Gennaro Chierchia and Sally McConnell-Ginet's (1990) introductory textbook in formal semantics. Consider first how they articulate their project in linguistics from the outset:

> The linguistic knowledge we seek to model, speakers' competence, must be distinguished from their observable linguistic behavior. Both the linguist and the physicist posit abstract theoretical entities that help explain the observed phenomena and predict further observations under specified conditions, (2).

Possible worlds are precisely the sort of "abstract theoretical entities that help explain the observed phenomena and predict further observations" that Chierchia and McConnell-Ginet have in mind here. When they end up introducing these entities, they reaffirmingly say, with explicit reference to the Stalnaker passage quoted above,

> The human activities on which the existence of possible worlds depends (through which they are stipulated) include using language and interpreting expressions. Semantics, as we understand it, seeks to develop a model (or part of a model) of such activities, (207).

On the face of it, what they say here might seem to be in line with what they say in the quote above it about positing theoretical entities to model semantic competence and thereby explain linguistic activities. However, though the above two quotes may seem to be compatible, there is a tension here. One the one hand, they claim that possible worlds depend for their existence on the linguistic activities that are a manifestation of semantic competence, "using language and interpreting expressions." On the other hand, they claim that these linguistic activities are to be explained by a semantic theory that features possible worlds. One cannot coherently maintain both of these claims at once.

Consider again the example from our toy language. The set of activities that are a manifestation of our speakers' semantic competence includes their acting in such a way that shows that they understand the sentences "*a* is gray" and "*a* is white" to be incompatible. They never utter both sentences at the same time, they correct speakers that do, and so on. These activities, we theorize, are manifestations of their knowledge of the meanings of the sentences "*a* is gray" and "*a* is white." Now, suppose we posit a class of theoretical entities that we call "possible worlds" in order to say what it is in which their knowledge consists. We say that their knowledge consists in their grasp of a particular fact about two sets of possible worlds, the ones that we theorize to be, due to their linguistic conventions, the correspondents of "*a* is gray" and "*a* is white." It is in virtue of knowing that these two sets are disjoint and knowing the conventions of their language that they know that one cannot correctly utter both sentences at the same time, and this knowledge explains why they act as we do, never uttering both sentences at the same time, correcting others that do, and so on. In this way, we explain their activity by taking them to bear a

cognitive relation to two sets of entities of the sort that we've posited—we explain their activity by saying that they know a certain fact about two sets of possible worlds, the set of worlds in which *a* is gray and the set of worlds in which *a* is white. If this is the form of our explanation of their linguistic activities, the existence of these possible worlds can't depend on the activities that they are posited to help explain. If they were so dependent, they wouldn't be able to figure into the explanation of these activities in the way that they do. So, if we take a possible worlds semantics to be able to give an account of the knowledge that they have in knowing the meanings of the sentences of their language, and we take this knowledge to explain their linguistic activities, we can't take the existence of possible worlds to depend on these activities.

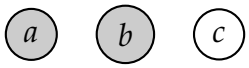
Once we drop Stalnaker's suggestion that possible worlds depend on the linguistic activities of speakers, I take it that it does not matter much whether we follow Lewis in taking possible worlds to be objects like the actual world or follow Stalnaker in taking them to be abstract properties like the property that the actual world instantiates in being just the way it is. Whatever we say—whether we say that other possible worlds are "other things" like the actual world or "other ways" like the way the actual world is—the basic picture is roughly the same: there is a space of primitive possibilities whose existence does not depend on us, to which we have cognitive access, and this cognitive access is a precondition for the possession of contentful mental states and semantic knowledge. This picture, whether Lewisian or Stalnakerian, is what I'll call the "primitivist picture." For the purposes of the present discussion, I will simply grant the cognitive access to worlds that we are supposed to have on the primitivist picture. Much ink has been spilled over the problem of our cognitive access to other possible worlds.[6] For my purposes here, I will grant—at least provisionally—that we have this access. The claim that I am about to make is that, even with this access to the myriad *particular* worlds being granted, the cognitive access to the world-involving *structures* needed for an extra-worldly semantic theory commits one to a fatal instance of the Myth of the Given.

[6]See O'Leary-Hawthorne (1996) for an overview.

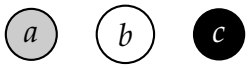
2.6 The Myth of the Extra-Worldly Given

The problem I am about to raise for extra-worldly semantics can be raised with respect to both propositions and properties. I will raise it just for the extra-worldly conception of properties here, since, given the compositionality of meaning, this suffices to raise a problem for the extra-worldly conception of propositions, and, as a result, for the whole theory. Let me first restate the extra-worldly conception of properties, which either are or are modeled by the semantic values assigned to (1-place) predicates.

On the extra-worldly conception of meaning, to grasp the meaning of a predicate is to grasp the property expressed by that predicate, where this property, modeled as a function from worlds to extensions, is thought of as something such that the grasp of it enables you to relate any possible world to a particular set of things in that world. For instance, the property of being gray is something such that grasping it enables you to take any possible world and single out the set of things that are gray in that world. So, if one grasps the property of being gray, then, given the following possible world:



one is able to single out the set $\{a, b\}$. Given the following possible world:



one is able to single out the set $\{a\}$, and so on for every possible world. Given that properties serve this cognitive role, they can be modeled as functions that map possible worlds to extensions. To have a grip on such a function would be to have a grip on exactly the sort of thing that would enable you to go from a possible world to the set of things that instantiate the property in that world. The theoretical role of properties, modeled by such functions, is that our grip on them explains this ability, no matter how things are, to sort things into the things that instantiate them and the things that do not. That's the thought.

The thought may all seem well and good, but I take it that there is a serious problem with it. I'll make this problem explicit by way of the following argument:[7]

1. One's grasp of a property can be adequately modeled by one's grasp of a function from worlds to extensions only if one's grasp of that property is identified with one's grasp of a rule/mapping that takes one from worlds to extensions, (premise).
2. One's grasp of a property explanatorily grounds one's grasp of a rule/mapping that takes one from worlds to extensions, (premise).
3. If A explanatorily grounds B, A cannot be identified with B, (premise).
4. So, one's grasp of a property cannot be identified with one's grasp of a rule/mapping that takes one from worlds extensions, (2, 3).
5. So, one's grasp of a property cannot be adequately modeled by one's grasp of a function from worlds to extensions (1,4).

This argument is clearly valid. To see whether it is sound, let's go through it premise by premise.

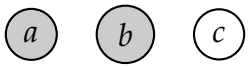
First, let us consider (1). Proponents of extra-worldly semantics will often speak of grasping a meaning as grasping a "rule" taking one from worlds to extensions. I write "rule/mapping" in (1) because the notion of a "rule" that is employed in this context can be nothing other than a (perhaps normatively expressed) mapping, something that takes one from a given world to an extension in that world.[8] A "rule," in this sense, is something that "tells one" (either literally or metaphorically, but presumably metaphorically): In case A, do x; in case B, do y; and so on. When Stalnaker (1976), for instance, speaks of understanding a sentence, grasping the proposition expressed by it, as knowing "the rule for determining the truth value of what was said, given the facts" (80)¡ he cannot be speaking of anything other than knowing a mapping from the various possible situations to a truth value. If this were not what he was speaking of, there would be no reason to think, on the basis of this explication, that grasping a proposition could be thought

[7]Note, when I say "property" in the context of this argument, I mean to be speaking of *simple material* properties, like the property of being gray.

[8]Note here that there is a difference between the intuitive notion of a "mapping" that I am using to explicate the use of the term "rule" here and the technical notion of a *function*. Following standard practice in formal semantics, I use "function" in its mathematical sense: a function f is a set of ordered pairs such that, for any x, y, and z, if $(x, y) \in f$ and $(x, z) \in f$, then $y = z$. A rule or mapping taking one from possible worlds to sets of objects is *modeled by* a set of ordered pairs that has, as first elements, possible worlds, and, as second elements, sets of objects.

of as the grasp of a function from worlds to truth values. The same point applies to the case with which we are presently concerned. In order to think that one's grasp of a property is adequately modeled by one's grasp of a function from worlds to extensions, one must think that what one grasps, in grasping a property, is a rule for determining the set of things that instantiate that property, given the facts, where a "rule" here, is simply a mapping. That is how we've been talking this whole time, and this way of speaking reflects the way of thinking that underlies this way of trying to model the meanings of predicates.

Now let us consider (2), which is the crucial, though I believe obvious, premise. Note first that, once again, I am simply assuming that we have cognitive access to possible worlds. We are, somehow, capable of getting various possible worlds "into view," in whatever, presumably metaphorical, sense in which we are supposed to be able to have them "in view." The question is: *how is it* that we are able to take any possible world that we have "in view" and single out the set of gray things in that world? Grasp of a mapping, of course, *does* enable us to go from any possible world to the set of things that are gray in that world, for, given any possible world, such a mapping simply "tells us," which things are gray in that world. But to answer the question of how it is that we are able to take any possible world and single out the set of gray things in that world by saying "We grasp a mapping taking any possible world to the set of gray things in that world" is not to answer the question at all. It is, in response to the question of *how* we are able to take any possible world and single out the set of gray things in that world, to say *we just are* able to do this. But there *is* an answer to the question of how it is that we are able to take any possible world and single out the set of gray things in that world. Indeed, there is an *obvious* answer. The answer to the question of how it is that we are able to take any possible world and single out the set of gray things in that world is that we know what it is for something to be gray and are capable of recognizing things as being such. That is, we grasp the property of being gray and, given any possible world that we have "in view," we are capable of ascribing it to gray things in that world. For instance, given the world



we are able to single out the set $\{a, b\}$ because we know what it is for something to be gray and we are capable of recognizing a and b as being such. That is, we grasp the property of being gray and we are capable of recognizing a and b as instantiating this property. This is obvious, but the possible world theorist is unable to say it. They are unable to say it because, from their point of view, it has things backwards, appealing to our grasp of the property of being gray in order to explain our grasp on a mapping from worlds to the sets of gray things in those worlds. If our grasp of the property of being gray is taken to *just be* this mapping, we cannot appeal to the former grasp to explain the latter.

Substantiating this last claim brings us to premise (3), which will be familiar from our discussion of the revised definition of possible worlds. I take this schema to be a statement of what is fallacious in the Euthyphro fallacy, the relevance of which for semantics has been brought to attention by Jason Bridges (2006).[9] When asked what it is for someone to be pious, Euthyphro claims that for someone to be pious is for them to be an object of God's love. However, when asked why it is that God loves the people that he does, Euthyphro claims that it is because they are pious. This pair of claims is inconsistent. If someone's being pious *explains* their being an object of God's affection, their being pious cannot be *identified with* their being an object of God's affection. (3) is a generalization of this fact, schematizing the explanans and explanandum that are not to be identified as A and B respectively. It's not at all clear that (3) needs or is capable of being given a justification from principles clearer than itself, but, if it does help, we can note that the truth of (3) follows from the fact that explanation is an asymmetric relation, whereas identity, if it is a relation at all, is clearly a symmetric one.

(4) follows from (2) and (3), and (5) follows from (1) and (4).[10] I conclude that an extra-

[9]Bridges focuses on how this fallacy plagues certain causal/informational conceptions of semantics, not considered here, but the critique is structurally similar.

[10]I take it that the logical structure here is quite transparent, and I'd be extremely surprised if any response to this argument involved rejecting any of the logical principles used in drawing this conclusion. For completeness, however, the first step is a universal instantiation, and the second step is either a modus ponens or tollens, depending on whether the "only if" specification of a conditional is formalized as a contrapositive or not.

worldly semantics is incapable of adequately modeling our knowledge of the meanings of predicates.

It is clear what the issue is here. One's grasp of properties explains one's ability to grasp functions going from worlds to extensions. Indeed, one's grasp of properties explains one's ability to grasp possible worlds at all. Possible worlds can only figure into semantic values on which one is supposed to have a grip insofar as these worlds are populated with objects whose properties one is capable of recognizing and ascribing to them. Only if one is capable of recognizing and ascribing properties can one have a grip on the mathematical objects with which the extra-worldly semanticist identifies properties. One must thus draw on one's capacity to recognize things as instantiating certain properties to ascribe the properties to them in order to have in view possible worlds in terms of which this capacity is supposed to be analyzed. In the words of McDowell, one must suppose that the availability of possible worlds for cognition does not draw on the capacities that are in fact required for this cognition.

Ultimately, I take it that the extra-worldly semanticist, insofar as they maintain their commitment to extra-worldly semantics as an explanatory enterprise, will be forced to accept one of two horns of a dilemma concerning our knowledge of extra-worldly semantic values: either our knowledge of them is *unintelligible* or it is *incoherent*. If one refuses to recognize the capacities that the grasp of semantic values in fact requires, then the resultant theory is one in which our grasp on this structure is unintelligible; we are simply taken to have this grasp without there being any explanation of how we do. If one does recognize the capacities that this grasp in fact requires, then the resultant theory is incoherent; this grasp is taken to not require the capacities that it is acknowledged that it does, in fact, require. This, recall, is the basic dilemma faced by anyone who is committed to an instance of the Myth of the Given. Extra-worldly semantics, as I've spelled it out here, is clearly an instance of the Myth. Knowledge structure of the world—understood here in terms of the set-theoretic organization of the possible worlds that populate the semantic universe—is taken to be simply given in such a way that precludes us from being able, even in principle, to understand this knowledge as rational, and so, as genuinely knowledge.

2.7 The Problem Percolates Up

This basic problem critically infects the whole extra-worldly semantic theory. Consider first (F3), the fact "a is gray" is incompatible with "a is white," considered originally in Section 1.3 as the sort of fact about our toy language for which we are supposed to provide an explanation. The explanation of this fact, on this semantic theory, is that the predicates "is gray" and "is white" are incompatible, and this theory purports to give us a way of modeling this incompatibility. If properties are modeled as functions from worlds to extensions, then we can also explain our grip on relations of incompatibility and entailment among properties in terms of our grip on the functions with which properties are identified. Two properties are incompatible just in case, for any possible world, their extensions are disjoint. So, given the first world shown above, one's grasp of the property of being gray enables one to single out the set $\{a, b\}$ and one's grasp of the property of being white enables one to single out the set $\{c\}$, and, given the second world shown above, one's grasp of the property of being gray enables one to single out the set $\{a\}$ and one's grasp of the property of being white enables one to single out the set $\{b\}$. Clearly, these two sets are disjoint, and so it is for any possible world. As such, the two properties are incompatible. Our grasp of the incompatibility between these two properties is modeled as our grasp of two functions such that, whenever they are given the same world as an input value, the sets of objects that are their output values are disjoint. However, if we can't model our grip on a property as our grip on a function from worlds to extensions, then we can't model our grip on the incompatibility of two properties in terms of our grip on two such functions. We can't, so the explanation of (F3) offered by our extra-worldly semantics goes out the window.

Now, as we've said, an extra-worldly semanticist will often not pay too much theoretical attention to the way in which the semantic theory assigns semantic value to content words, and so may, at this point, simply say that they are not concerned with explaining facts such as (F3). Consider, however, a set of facts for which an extra-worldly semanticist generally *will* want to provide an explanation. The sentence of "a is gray and b is white" entails the sentence "a is darker than b," but not vice versa. However, if you stick

a negation operator in front of these two sentences, the entailment relation goes in the other direction: the sentence "It's not the case that a is darker than b" entails the sentence "It's not the case that (a is gray and b is white)," but not vice versa. One might take a possible worlds semantic framework to be able to explain this fact. In such a framework, atomic sentences are assigned sets of worlds as semantic values and "and" expresses the operation of taking the intersection of two sets of worlds. The set of worlds assigned to "a is gray" will be the set of worlds in which a is gray, the set of worlds assigned to "b is white" will be the set of worlds in which b is white, and so the set of worlds assigned to "a is gray and b is white" will be the intersection of these two set of worlds. The set of worlds assigned to "a is darker than b" will be the set of worlds in which a is darker than b. As a matter of fact, the first set is a subset of the second one. So, thinking of the entailment relation in terms of the subset relation, "a is gray and b is white" entails "a is darker than b." Insofar as there are some worlds in which a is darker than b but it is not the case that a is gray and b is white (for instance, any world in which a is black and b is gray), the converse is not true. Now, by taking "It's not the case that" to express complementation, it seems that we can explain why, when we embed these sentences under this negation operator, the entailments are reversed. Complementation reverses the subset/superset relation between sets: if $A \subseteq B$ then $C - B \subseteq C - A$. So, by understanding entailment in terms of the subset relation and understanding negation in terms of complementation, we can explain why the entailment relations are reversed when the sentences are negated, and, by taking speakers' knowledge of meaning to be (adequately modeled by) knowledge of these semantic values, we can explain their knowledge of this fact. If I am right, however, there is no explanation to be had here.

The explanandum here—the fact to be explained—is that the entailment relations between sentences are reversed when these sentences are negated. The explanans—the set of facts doing the explaining—is (1) that sentences have sets of possible worlds as semantic values, (2) that entailment between sentences is a matter of the subset relation obtaining between their semantic values, and (3) that a negation operator semantically expresses the operation of complementation. I take the argument that I have just given to have undermined (1), and, in doing so, undermined the whole explanation. Since the

theory is compositional, such that the semantic value of a sentence must be composed of the semantic values of its parts, then, if functions from worlds to extensions cannot serve as semantic values of predicates, then sets of possible worlds cannot serve as the semantic values of sentences. If sets of possible worlds cannot serve as the semantic values of sentences, and so we cannot think of entailment in terms of the subset relation, nor can we think of negation in terms of complementation. If we can't think of entailment in terms of the subset relation and negation in terms of complementation, we have no explanation for the fact to be explained. So, if the argument I have just given goes through, then possible worlds semantics can function as no more than an analytical tool for elucidating facts such as the fact that entailment relations between sentences are reversed when these sentences are negated. We can use possible worlds semantics as a tool for systematically bringing into view the set of semantic relations that obtain between sentences of the language for which we are giving a semantic theory.[11] However, we cannot, with the use of such a framework, explain what it is in virtue of which these semantic relations obtain. Insofar as the aim of a semantic theory is not merely elucidatory but explanatory, possible worlds semantics fails as a semantic theory; possible worlds semantics can play a role in an explanatory semantic theory, but only if it is supplemented by a semantic theory that is actually able to do the explanatory work that it cannot.

2.8 Conclusion

While possible worlds semantics may be a fine tool for systematizing a set of semantic facts of which we already have knowledge, we are not going to get an account of these facts or our knowledge of them through the use of a possible worlds semantics. Possible worlds semantics is not going to give us an account of linguistic meaning or our knowledge thereof. For such an account, we must look towards a different sort of semantic theory. In the next chapter, I'll consider an alternative sort of worldly semantic theory—intra-worldly semantics—focusing on the variant of such a semantic theory proposed by Scott Soames and Jeff King.

[11] In general, we can use model-theoretic tools to elucidate inferential relations, and possible worlds semantics for natural language is one instance of this general fact.

3

Intra-Worldly Semantics

3.1 Introduction

In this chapter, I'll consider semantic theories that take the form of what I'll call "intra-worldly semantics." In the paradigmatic case, such a semantic theory will be one in which we think of the meaning of a sentence as a structured proposition which ascribes properties or relations to objects, representing objects as instantiating properties or standing in relations. As a variant of worldly semantics, an intra-worldly semantics requires us to try to comprehend our knowledge of meaning sentences and predicates as assymetrically dependent on our knowledge worldly entities and their relations. Specifically, an intra-worldly semantics thinks of this knowledge as knowledge of objects, properties, relations, and primitive modal relations between properties and relations. I will argue that we cannot comprehend our knowledge of meaning in this way. The core problem, which will be familiar in its basic form, is that the intra-worldly knowledge to which an intra-worldly semantics appeals can only be understood *in terms of* our knowledge of the semantic rules governing the correct use of predicates, but this knowledge of semantic rules, on an intra-wordly semantics, is understood as *depending on* our knowledge of primitive modal relations between properties and relations.

3.2 The New Non-Primitivist Actualism

Let us introduce intra-worldly semantics by returning to the issue of defining possible worlds. In the previous chapter, I considered two "primitivist" views of possible worlds, the genuine realism of David Lewis and the "modest realism" of Robert Stalnaker. The

key difference between these two views is that, whereas Lewis takes possible worlds to be genuine other worlds of exactly the same sort as the actual one, Stalnaker takes possible worlds to be properties—ways for the world as a whole to be—that exist in the actual world. In my eyes, however, when compared to the boarder class of views about possible worlds, the key similarity between these views is greater than the difference. What is crucial to both Lewis and Stalnaker's accounts is that neither try to define what possible worlds are. At least for the purposes of the semantic theory, they take possible worlds to be basic. This contrasts with views that try to give a constitutive account of possible worlds, saying what they are in terms of more basic entities such as states of affairs, propositions, or properties. These are non-primitivist views: views according to which possible worlds are taken to be constituted by and definable in terms of more ontologically basic entities. Often, non-primitivists align themselves with Stalnaker as "actualists," but the real crucial divide, in my eyes, is between the primitivists, like Lewis and Stalnaker, and the non-primitivists such as Robert Adams (1974) and Alvin Plantinga (1976), of the old days, and Jeff King (2007a) and Scott Soame (2010), of the new days. Thinking of possible worlds as constructed from more metaphysically basic entities leads to a very different picture of meaning than the extra-worldly one.

I'll focus here on the new non-primitivist actualism, which has emerged most prominently in the work of Jeff King (2007a) and Scott Soames (2010). On this view, possible worlds are "big uninstantiated properties that are complex and have as parts other properties and relations," (King 2007a, 447). That is, they are complex properties that the actual world could have instantiated (and would have instantiated had things been otherwise), that have, as constituents, other simpler properties and relations. To get this conception of possible worlds into view, first consider the thought that there can be the properties that the world as a whole might instantiate—for instance, the property of being such that a is black. If a is black, the world as a whole is such that a is black. That is, it instantiates the property of being such that a is black.[1] Now, consider with the thought that properties can be joined together to form conjunctive properties. For instance, the property of

[1]For Soames (2010), the relevant property is spoken of as the property of making the proposition that a is black true, but this may plausibly be taken to be the very same property as the property of being such that a is black. In any case, the details don't matter for our purposes.

being white and the property of being round can be conjoined to form the conjunctive property of being white and round. Since there are properties that the world as a whole might instantiate and properties can be conjoined to form complex properties, there is, for instance, the property of being such that *a* is black, *b* is white, and *c* is gray. This is a property that the world could have instantiated and would have instantiated if it were actually such that *a* is black, *b* is white, and *c* is gray. This property is a way the world could have been, a possible world, or, as we should say if our terminology is not to be misleading, a "possible world-state."

King and Soames, taking possible world-states to be complex properties that the world could have instantiated, understood in this way, both say that, in addition to possible world-states, there are impossible world-states, complex properties of the same sort that the world could not have instantiated. It's not hard to see why this is a natural conclusion to draw on a view of this sort. As we've already said, properties can be conjoined to form conjunctive properties. We've also said that there is a property of being such that *a* is black, and there is a property of being such that *a* is white. So why shouldn't we think that there is the conjunctive property of being such that *a* is black and such that *a* is white? As King (2007a) says, "if you hold that properties exist, but deny that properties of a certain sort exist, you should provide a principled reason why properties of that sort don't exist," (448), and it's hard to see what our principled reason could be here. If there is this property, then clearly there is a property of being such that *a* is black, *a* is white, *b* is white, and *c* is gray. If we think that possible worlds are properties of this sort, what could that property be other than an impossible world? King follows this thought through and claims that impossible worlds exist. It's just that, unlike (non-actual) possible world-states, which are only *contingently* uninstantiated, impossible world-states are *necessarily* uninstantiated. Whereas possible world-states are ways that the world could have been, impossible world-states are ways that the world could not have been.

It is important for primitivists about possible worlds such as Lewis and Stalnaker to maintain that there are no worlds that are impossible, for Lewis and Stalnaker want to understand what it is for some state of affairs to be impossible in terms of the fact that there is no world in which it obtains. If there are impossible worlds, no analysis of this sort can

be maintained. However, King and Soames do not wish to provide a reductive analysis of modal notions in terms of possible worlds. According to Soames, one of Lewis's main errors consists in his "thinking that modal notions can be analyzed away, rather than taken as primitive," (2010, 110). If we take modal relations that obtain between properties as primitive, then we can demarcate the set of possible worlds from the set of impossible ones in terms of whether or not it's possible for the world to be instantiated. Cashing this out, if we think possible worlds are complex properties that have simpler properties as their constituents, we can demarcate the possible world-states from the impossible ones by specifying whether simple properties that cannot possibly be co-instantiated by some object would have to be co-instantiated by some object in order for a world-state to be instantiated by the world as a whole. By taking the compatibility or incompatibility of simple properties to be explanatorily prior to the possibility or impossibility of world-states in this way, we can demarcate the possible from the impossible world-states. So, since the property of being black is incompatible with the property of being white in the sense that the two properties cannot possibly be co-instantiated by a single object, the world cannot instantiate the property of being such that *a* is black and being such that *a* is white; instantiating this property would require a single object (namely, *a*) to instantiate both the property of being black and the property of being white, and that is not possible. So, any world-state that includes the property of being such that *a* is white and being such that *a* is black is not a possible world-state.

The basic metaphysical idea here, underlying the new non-primitivist actualism, is, as Michael Jubien (2008) puts it, that "relations among properties are the real source of our intuitions about necessity and possibility," (104). Spelling out this account, Jubien tells us that we can provide the following explanation for the fact that, if something's a horse, then it must be an animal, or to use our example, that something black (all over) cannot be white (all over):

> *It's that the two properties' intrinsic natures together guarantee it.* We may therefore see this connection as an 'intrinsic relation'—one that holds between the two properties strictly as a result of their individual intrinsic natures. Here is the locus of the needed 'modal oomph'. Differences between properties' own intrinsic properties establish modal connections between them, (93).

So, on Jubien's account, modal facts obtain in virtue of the "intrinsic natures" of properties these facts involve. For instance, the fact that, if *a* is black, then it can't be the case that *a* is white is explained by the intrinsic natures of the properties of being black and being white which jointly establish the modal relation of incompatibility obtains between them. In providing a similar sort of account of necessity and possibility, Bob Hale (2013) appeals to the "essences" of properties—*what it is to be* these properties—in order to explain modal facts. Hale tells us:

> No matter what entity or kind of entity is in question—be it a kind of object, or property, or relation, or function, or thing of some other kind—there will be some facts about what it is to be that entity (or an entity of that kind), and these will give rise to corresponding necessities, (147).

So, on Hale's preferred way of putting things, the fact if *a* is black, then it can't be the case that *a* is white is explained what it is to be black and what it is to be white, where part of what it is to be black is to be non-white. However one prefers to spell out the details, the basic idea, in both of these accounts, is that the properties, in virtue of their natures or essences, bear certain modal relations to one another, and it is in virtue of these modal relations, and our grip on them, that we are able to say what states of the world are possible or not.

3.3 A Simple Intra-Worldly Semantics

Thinking of possible worlds in this way lends itself to a very different sort of semantic theory. If one explains which worlds are possible by appealing to relations of compatibility and incompatibility between properties, one can't then turn around and analyze these relations in an extra-worldly framework. Rather, properties and their modal relations are taken as primitive, for the purposes of the semantic theory. We thus get a different kind of semantic theory—an intra-worldly semantics. Both King and Soames propose versions of an intra-worldly semantics. The details of their theories differ, but those differences don't matter much for our purposes here. For King, names are assigned objects as semantic values, 1-place predicates are assigned properties, and *n*-place predicates are assigned *n*-place relations. Sentences are assigned propositions, which are composed

out of these semantic values and represent objects as instantiating properties or standing in relations.[2] Soames assigns names, 1-place predicates, and n-place predicates, *acts of cognizing* objects, properties, and relations as semantic values, and takes propositions to be acts of predicating cognized properties and relations of cognized objects, but everything basically works out the same. For simplicity, I will consider a view along the lines of that King proposes in which names are assigned objects and predicates are assigned properties and relations.

How does a semantic theory that appeals to intra-worldly facts of this sort work? At the most basic level, such a semantic theory assigns objects to names, properties to 1-place predicates, and n-place relations to n-place predicates. For our simple toy language, for example, the assignment of basic semantic values might be the following:

$[\![\textbf{a}]\!] = a$	$[\![\textbf{is black}]\!]$ = the property of being black
$[\![\textbf{b}]\!] = b$	$[\![\textbf{is gray}]\!]$ = the property of being gray
$[\![\textbf{c}]\!] = c$	$[\![\textbf{is white}]\!]$ = the property of being white

$[\![\textbf{is darker than}]\!]$ = the relation of being darker than
$[\![\textbf{is lighter than}]\!]$ = the relation of being darker than
$[\![\textbf{is the same color as}]\!]$ = the relation of being the same color as

The semantic value of a sentence is a proposition. A sentence consisting in a name concatenated with a 1-place predicate expresses a proposition that represents the object that is the semantic value of that name as instantiating the property that is the semantic value of that predicate. A proposition that represents an object as instantiating a property is true if that object instantiates that property, false if that object does not instantiate that property. For instance, the sentence "*a* is gray" expresses the proposition that *a* is gray, which represents *a*, the semantic value of "*a*," as instantiating the property of being gray, the semantic value of "is gray," and so is true just in case *a* is gray.

For logical operators, different intra-worldly semanticists have somewhat different proposals, and the details of any such proposal don't particularly matter for our purposes

[2] King's account of just what the relation is that binds together an object and a property in a proposition such that the proposition represents that object as instantiating that property is very complicated, and, in my view, utterly confused. For a criticism, see Simonelli (M.S.f). For our purposes here, the details of the theory don't matter.

here. One simple proposal, suggested by King (2009, 114 n. 28) and also adopted by Hanks (2011), is to take the semantic values of the logical operators to also be properties or relations, but ones that are ascribed to and instantiated by propositions rather than objects. So, we might assign to the logical operators the following properties and relations instantiatable by propositions:

⟦It's not the case that⟧ = the property of being false.
⟦and⟧ = the relation of both being true
⟦or⟧ = the relation of at least one being true

So, the proposition "It is not the case that *a* is gray" expresses the proposition that it is not the case that *a* is gray, which represents the proposition that *a* is gray, the semantic value of "*a* is gray," as instantiating the property of being false, the semantic value of "It is not the case that," and so true just in case it is not the case that *a* is gray. And so on. We thus have a simple compositional semantic theory for our toy language which appeals only to entities in the world—objects, properties, and relations—and no worlds as a whole.

Assigning names objects as semantic values and 1-place predicates properties as semantic values enables us, according to King, to "give a simple, direct explanation" (785) of facts such as those consisting of the "robust judgments about entailment relations between sentences" (784) that speakers make. Consider, for instance (F2), the fact that the sentence "*a* is black" and the sentence "*b* is gray" jointly entail the sentence "*a* is darker than *b*." An intra-worldly semantics of the sort proposed by King takes this fact to be explained in part by the natures of the following three entities in the world: the property of being black, the property of being white, and the relation of being darker than. These three entities stand in a certain relation: if something instantiates the property of being black, and something else instantiates the property of being gray, then the first thing stands in the relation of being darker than to the second thing. The fact that these three entities stand in this relation is not a semantic fact but a worldly fact; it is a fact consisting in three entities in the world (two properties and a relation) standing in a certain relation. Furthermore, the fact has a certain sort of modal robustness. It doesn't just happen to be the case that, if one thing is black and another thing is gray, then the first thing is darker than the second thing. Rather, if one thing is black and another thing is gray, then

the first thing *must* be darker than the second thing. Following the line taken by Jubien (2008) and Hale (2013), the modal robustness of this fact is grounded in the essences of the properties and relations it involves. It follows from what it is for something to be black, what it is for something to be gray, and what it is for one thing to be darker than another that, if something is black, and something else is gray, the first thing is darker than the second. From the fact that the property of being black, the property of being gray, and the relation of being darker than stand in the relation specified above, and the fact that the expressions "is black," "is gray," and "is darker than," have these properties and relations as semantic values, it follows that the sentences "*a* is black" and "*b* is gray" jointly entail the sentence "*a* is darker than *b*." It seems, then, that our simple intra-worldly semantics gives us a "simple, direct explanation" of the behavior we set out to explain, just as that consisting of "judgments about entailment relations between [atomic] sentences." Once again, however, things are not how they seem.

The intra-worldly semanticist appeals to properties at the base level of their semantic theory. Accordingly, it is these entities, and our grasp of them, that needs to be investigated if we are to investigate the foundation of an intra-worldly semantics. What are these entities and how do we grasp them?

3.4 Properties, Appealed to and Unnacounted for

Though many semanticists appeal to properties at some level in their semantic theory, most don't take them to be semantic primitives. If you ask a working semanticist what a property is, they're likely to answer this question in an extra-worldly framework, saying, for instance, that a property is a function mapping each possible world to a set of objects. This is, of course, a definition of a property as mathematically constructed from more primitive entities—objects and possible worlds, which are taken as basic from the point of view of the semantic theory. It is not a definition of properties that takes properties themselves as basic. In contrast to the extra-worldly semanticist, the intra-worldly semanticist does not want to think of properties as constructed from objects and possible worlds. This may well be because they rightly see that the particular objects that populate

those possible worlds cannot be understood as the distinctive objects that they are independently of the properties they instantiate, and possible worlds cannot be understood as genuinely *possible* rather than *im*possible apart from thinking about the properties that would have to be co-instantiated by objects if some world were actual. So properties can be no less conceptually basic than objects and, indeed, must be more conceptually basic than possible worlds. But what *are* properties? They are, of course, the things that serve as the semantic values of predicates in the semantic theory. So, they are what speakers grasp in grasping the meaning of a predicate. But, once again, what *do* speakers grasp in grasping the meaning of a predicate? The intra-worldly semanticist, insofar as they are proposing an account of speakers' knowledge of meaning, ought to have *something* to say in response to this question, so they ought to be able to say something about the properties that are theorized to be the semantic values of predicates.

There are really two questions here. The first is what are properties *in general*? That is, what constitutes the ontological category of properties? This first question, I don't think, is too difficult. Consider first Van Inwagen's (2006) identification of the notion of a property with the notion of a "thing that can be said of something," (27). According to Van Inwagen, properties, relations, and propositions all belong to the same fundamental type: they are all assertables. Properties are assertables with one place for a thing for the property to be asserted of, n-place relations are assertables with n-places for things for the relation to be asserted of, and propositions are assertables with no place for any things to fill—so, they are asserted not *of* anything, but *full-stop*. Things that can be asserted *of* things, whether just one or many, are *predicables* or *ascribables*. So properties, on this account, are things that can be predicated of or ascribed to things, said of them. This is one quite common construal of what properties are. Consider now Hale's (2013) construal of what properties are:

> A property, on this account, just is a condition which things may or may not satisfy. [. . .] A property, we might say, is a way for things to be—perhaps a way some things are or could be, but perhaps a way nothing could be, (165-166).

This is, in fact, how Soames (2015b) answers the question of what properties are, when pressed on the issue, and this is also quite a common construal. Bringing these two con-

ceptions together I think there is a relatively philosophically unobjectionable specification of what, in general, properties are: they are ascribables and instantiateables. That is to say, they are things that can be both ascribed *to* objects and instantiated *by* objects. This is, indeed, how they are characterized in the opening paragraph of the *Stanford Encyclopedia* entry on properties (Orilla and Paolini Paoletti 2020), and, of course, this is just what they need to be be if they are to play the role in an intra-worldly semantic theory that they are supposed to play. If a proposition ascribes a certain property to a certain object, then that proposition is true just in case that object instantiates that property, false if that object doesn't instantiate that property.[3]

What else should we say of properties in general? Theorists like King and Soames, who appeal to properties at the base-level of their semantic theories, typically say very little.[4] However, we can find at least a few remarks in King's work about properties, further specifying what they are. King's clear that he thinks of them as "entities in the external world" (2018, 784) "existing quite independently of minds and languages" (2007, 450), that some of them are complex in the sense of being "made up of other properties and relations" (1998, 157; 2007, 447), and that they stand in relations of entailment to other properties (1998, 173 n8). So, though King never explicitly provides a general answer to the question of what properties are, he is happy to specify that they have just the features he needs them to have in order for them to play just the role he wants them to play in the sort of explanation of semantic competence that he wants to give, serving as the contents of predicates. We might reasonably immediately wonder whether this sort of move is really justifiable. One might be inclined to liken it to postulation in other areas of scientific inquiry. For instance, of course, in the course of doing astronomy, we may posit an object with a certain gravitational force to make sense of the observed behavior of other celestial objects. One might think of properties as theoretical posits along the

[3]Note that, for Soames, a proposition is an act type that ascribes a property to an object only in a derivative sense that an agent that tokens that act in thought or speech ascribes that property to an object.

[4]One notable exception is Peter Hanks (2015, 2017) who, though he appeals to properties in much the way that King and Soames do, worries about these appeals in the context of a semantic theory and thinks that we need some account of how our grip on properties is achieved such that we understand the "standards of correctness" they provide for acts of predication. Though the framing of the issue here is quite different than that of Hanks, and so I do not explicitly engage with him here, the positive account that will be provided in chapters four and five will answer these concerns though they do so only at the cost of the explanatory use to which Hanks hopes to put properties in the context of his semantic theory.

same lines.[5] However, the assumption implicit in a scientific postulation is that further scientific inquiry will eventually (or at least in principle *could*) tell us *what this thing is*, be it a gaseous planet, a large asteroid, or what have you. In this case, however, we don't even have so much as a gesture at what a proper account of these things would look like.

This brings us to our second question, which will prove more decisive for our project here: what are the *particular* properties that figure into the semantic theory? What, for instance, is the property of being black, appealed to in the intra-worldly semantics for our toy language? Or, to turn to a real language for a moment, what are the various properties that we would appeal to in providing an intra-worldly semantic theory for a natural language such as English? To get a sense of what an account of the properties that figure in an actual semantic theory would have to encompass, consider all the properties we would need to be able to appeal to in order for such a theory to be explanatory adequate. We need not just what one might regard as more fundamental properties, such as the property of being red, being green, being square, being round, and so on; being positively charged, negatively charged, and so on. We need *all* the properties corresponding to predicates of the language. So, we need the property of being a reptile, a bird, a cardinal, a penguin; the property of being a chair, being a table, a cup, a flask; the property of dancing, of swimming, of skiing, of smoking; the property of being a fruit, a vegetable, a steak, milk, coffee, tea, chocolate cake; the property of being a novel, a novella, a screenplay; and many *many* more. As competent speakers of English we conceptually grasp this *vast network* of properties and their inter-relations, each one the content of a predicate of the language we speak. Without any story about what the constituents of this vast realm of properties are, or how speakers come to have a grip on them, an intra-worldly semantic theory, based on the assumption that speakers *do* have a grip on them, hangs in the air.

Now, perhaps theorists like King and Soames simply think that, whatever story is to

[5]Indeed, Jubien (2008), whose work King (1998) cites on the nature of properties, explicitly makes this comparison, saying:

> Properties have central roles to play and we speak every day as if they are playing them right there. As long as there are no genuine problems with properties, we should welcome them as entirely sensible theoretical posits, (2008, 42-43).

be told about the structured space of properties on which we have a grip, there surely is *some* story to be told, and it's simply not their job to tell it. Perhaps it's the job of the metaphysicians. King and Soames don't explicitly say this, since they don't explicitly address this question (very few semantic theorists actually do), but it's presumably the sort of line that the intra-worldly semanticist would want to take. Perhaps the only pair of semanticists who do explicitly adress this question is Herman Cappellan and Ernie Lepore (2005), and they do take this line, so let us consider their explicit defense of it. To provide some context, Cappellen and Lepore endorse a "minimalist" semantic theory according to which, for instance, the sentence "*A* is tall" expresses the proposition that *A* is tall and is true just in case *A* is tall. On their account, as far as semantics is concerned, the analysis ends there. This is analogous, they claim, for how the sentence "*a* is red" expresses the proposition that *a* is red and is true just in case *a* is red or how the sentence "*A* dances" expresses the proposition that *A* dances and is true just in case *A* dances. Now, though their minimalist view is quite controversial in contemporary philosophy of language when applied to expressions like "tall," which are widely taken to be context-sensitive and so in need of some sort of further semantic analysis, the line they take with respect to "red" is supposed to be quite uncontroversial, and taking this line with respect to "dances" is supposed to be completely uncontroversial. Their argumentative strategy, then, is first to argue why it is patently unreasonable to demand that the semanticist give an account of what it is to be *red* or what it is to *dance*, and then to extend this reasoning to the question of what it is to be tall. They assume their reader will regard the sort of objection they're arguing against, concerning the properties of being red and dancing, as "borderline silly" (156), but the arguments they give in response to this "silly" objection concerning "red" and "dances" are supposed to extend to the philosophically respectable objection regarding "tall." My aim here, of course, is not to get into the debate within intra-worldly semantics concerning the correctness of minimalism or contextualism. Rather, it is to show that this "silly" objection, which would apply to both minimalism and contextualism, is really quite serious for the project of intra-worldly semantics as a whole.

The first point to make is that, insofar as these properties play an essential explanatory role in an intra-worldly semantic theory, these questions about *what they are* are certainly

legitimate ones. Division of labor is fine, but, somewhere along the line, *someone's* got to do the labor. At the very least, even if the labor is not *actually* going to get done for each case, we ought to be sure that it's *possible* to actually do it, and, moreover, have some idea of how it *would be* done *were* we to actually do it. One might think, then, that Cappellen and Lepore would have some remarks about how this division of labor is supposed to work, how the respective disciplines of semantics and metaphysics can ultimately be connected in a complete theory of linguistic and conceptual competence. Yet, when Cappellen and Lepore actually discuss the sort of metaphysical work that *would* give answers to these questions about the properties that figure in the sort of semantic theory they propose, they are clearly quite pessimistic regarding the possibility that it can be productively done or, indeed, done at all. Indeed, their rhetoric makes it quite clear that they don't take this sort of work seriously in the slightest. They tell us that, if you seriously ask these questions about what the properties that figure into the semantic theory are, "you'll regret it because it'll just turn into a rather large metaphysical mess (*not* a mess of our making, just the regular mess metaphysicians inevitably like to throw themselves into)" (158). The rhetoric here suggests not just that the sort of metaphysics that would seek to answer these questions is *hard*, like, say, topology is hard, but, rather, that it's *hopeless*. This dismissive rhetoric, however, is in tension with any real talk of a "division of labor" between semantics and metaphysics. Consider the following analogy. If one is proposing a biological theory, and, in response to some question concerning the things to which one appeals to in that theory, says that it's a matter of chemistry how those questions are to be answered, it's fine if one admits that chemistry is hard, perhaps even too hard for one to do oneself, but one should surely not say that chemistry is hopeless!

Now, of course, even if Cappellen and Lepore have this pessimistic attitude towards the metaphysical enterprise of providing an account of the properties that figure into an intra-worldly semantic theory, it need not be the case that every intra-worldly semanticist has this attitude. Perhaps intra-worldly theorists have or at least ought to have a more optimistic attitude towards the metaphysical enterprise that would provide in accounts of the properties that figure in the intra-worldly semantic theory they propose. Just to be clear, I am myself quite optimistic about this enterprise. Indeed, I will eventually

show that this enterprise is not hopeless at all, but in fact quite doable. However, what I'll now argue is that this enterprise is indeed hopeless, given the constraints put on it by the explanatory use to which properties are put in an intra-worldly semantic theory. There *is* an account of the properties that are grasped by speakers of a natural language to be given. By the end of this chapter, we will say just what that account is, and, by the end of the dissertation, we will have developed the key tools needed to actually fill it in. The problem, however, is that this account of properties is simply not available to a proponent of an intra-worldly semantic theory, given the explanatory use to which properties are put in such a theory. Before actually getting to this account and explaining its incompatibility with an intra-worldly semantic theory, let us first go through the failure of various proposals that might be considered viable from the perspective of intra-worldly semantics.

3.5 The Problem of Defining Properties

Sticking with the strategy of examining semantic theories by considering how they fare with respect to a very simple toy language, let us consider how someone who is proposing an intra-worldly semantics for our toy language might try to specify the meanings of one of the predicates belonging to it—the predicate "black," say. According to the intra-worldly semanticist, the predicate "black" expresses the property of being black. This property figures in at the base level of the semantic theory. How should one say what this property is? Let us consider some initial attempts.

Of course, it would be absolutely hopeless to try to define the property of being black as follows:

> $[\![\textbf{black}]\!]$ = the property of being black =
>
> the property that an object instantiates just in case that object is black.

This, of course, says nothing. It is indeed the case that the property of being black is the property that an object instantiates just in case it is black, but one should not be tempted to hear this as a substantive definition of what the property of being black is. What such a statement is, really, is just a substitution instance of a grammatical remark expressing how

property-talk in general is to be used. We can introduce property-talk into a language by specifying, among other things, that we are entitled to say (or committed to saying) "*a* instantiates the property of being *F*" just in case we are entitled to say (or committed to saying) "*a* is *F*." Given this schema, it is of course, true of the property of being black that it is instantiated by an object just in case that object is black, and it is likewise true for any property, but that's because it's a completely empty description, one that says absolutely nothing about any particular property at all. Accordingly, such descriptions can't function to specify what the particular properties that figure in at the base level of our intra-worldly semantic theory actually are. They are, quite literally, without content. Since what we are supposed to be specifying is, of course, the semantic *content* of the predicate "black," such a definition will not do.

A different way to try to specify what the property of being black is would be to try to do so along the following lines:

⟦black⟧ = the property of being black =

the property that all and only the black things instantiate.

Of course, unless we appeal to possible worlds, this isn't going to work. After all, it might just so happen that all and only the black things are spherical, say, rather than cubical. In such a case, this definition and the definition of "spherical" would specify exactly the same property, since the exact same things would instantiate them. Clearly, however, the property of being black and the property of being spherical are not the same property. So this definition is blind to the difference between properties that just happen to be instantiated by the same things. Now, one might try to get around the issue with possible worlds by adding a primitive necessity operator, transforming the above definition into the following:

⟦black⟧ = the property of being black =

the property that, necessarily, all and only the black things instantiate.

However, trying to make any sense of the sort of necessity at play here without appealing to possible worlds simply brings us back to the first definition, where the necessity is

understood in terms of the fact that this statement follows simply from the grammar of property-talk. Similarly, one could add to the end of this definition "in virtue of being black," but, once again, this just brings us back to the first definition. So, any definition along these lines is not going to do.

Given the failures of the above two definitions, one might think that the answer to the question of what the property of being black is is not a conceptual question at all, but, rather, an empirical question, something to be answered by empirical investigation into the nature of black things. Suppose, upon conducting such an investigation, we come to the following conclusion:

> **[[black]]** = the property of being black =
>
> the property of completely absorbing light.

This does seem to give us a substantive specification of what the property of being black is. The obvious issue here, however, is that it doesn't give us a substantive specification of what the speakers of our toy language grasp in grasping the meaning of the predicate "black." The hypothetical speakers of our toy language, we may suppose, grasp the property of being black in grasping the meaning of the predicate "black" which belongs to their language without having any grip on the property of absorbing light. They have no words for the property of absorbing light and so we have no reason to think that they know what it is for something to absorb light at all. So, this specification of what the property of being black is cannot be a specification of what that speakers of our toy language grasp in grasping the meaning of the predicate "black."

Now, perhaps a defender of the above definition will want to say that it's *really* this property on which they have a grip on, even though they don't have have a grip on the essence of this property. Recall, however, the explanatory role of properties in the context of the semantic theories. Speakers are supposed to have *some* grip on the essences of these properties, since their grip on these essences is supposed to explain their grip on the modal relations that the properties bear to one another. Perhaps, if we take this line we can explain these modal relations *ourselves*. For instance, if we take this line with respect

to the property of being black, then, taking the same line with respect to the property of being white, we might define it as follows:

⟦white⟧ = the property of being white =

the property of completely reflecting light.

It may well be an aspect of our conceptual framework that nothing can both completely absorb and completely reflect light, and so we can account for *our* grasp of the incompatibility of these properties in this way. Our task as semantic theorists however, is to give an account of the grasp of the *speakers of the language for which we are constructing a semantic theory.* Any such definition, which is blind to the distinction between what is grasped by the speakers of the language for which we are giving an intra-worldly semantics and what we grasp ourselves as theorists, won't do.[6]

In response to the failure of these last three definitions, it might seem that the problem is the very idea that the property of being black can be captured in words. Perhaps, because of the particular sort of property that the property of being black is, a simple qualitative property, words will not do. If that's so, then perhaps the right way to specify the property of being black is to do so as follows:

⟦black⟧ = the property of being black =

the shade instantiated by the following object:

Here, one *shows* the reader the property of being black, by showing the reader an instance of it, rather than trying to *say* what it is. Upon being shown something that visibly instantiates of the property of being black, the reader is supposed to know the specific property that figures in the semantic theory by simply being shown that property. Now,

[6]Hale (2007) suggests this sort of answer, saying "If what is in question is being red as a property of surfaces (as distinct from the property of light, or the property of sense-impressions), *being coloured* consists in reflecting light in the visible spectrum (roughly 390–750 nm), and the 'more' is that what is red reflects light of wavelengths of roughly 630-740 nm," (147 n5). In the context of semantics, this fails for the reason specified here. In Chapter Six, we'll see that this even fails in the context of an attempt at scientific specification of the property of being red, understood as a theoretical property.

if one goes this route for the property of being black, then surely one would go the same route for the property of being gray:

⟦**gray**⟧ = the property of being gray =

the shade instantiated by the following object:

It might seem as if this is the way to go, at least for very simple properties such as color properties that don't seem like they can be constructed in any way from other properties.

As Wittgenstein (1953) pointed out, however, such "ostensive definitions" are not going to work. To see this, consider how one might attempt to ostensively define the relation of being darker than. This is supposed to be the semantic value of "is darker than" that figures in the semantic theory. Accordingly, we should be able to specify what it is as well. Attempting the same strategy here, however, yields obvious problems. Consider the attempt to try to define this relation as follows:

⟦**darker than**⟧ = the relation of being darker than =

the relation instantiated by the following two objects, with the one on the left occupying the first place of this relation and the one on the right occupying the second place in this relation:

The problem here, of course, is that there are indefinitely many relations instantiated by these two objects. For all that is said here, the demonstrated relation could be the relation of being to the left of, the relation of being the same shape as, or any one of a great number of relations. The demonstration itself does nothing to ensure that the reader takes it to be the relation of being darker than that is demonstrated rather than any one of a number of other relations that these two objects stand in. Now, presumably, you *did* take it to be the relation of being darker than that was demonstrated here rather than one of these other relations. But that's only because you read the text above the demonstration and,

knowing what "darker than" means, you knew it was the relation of being darker than that was supposed to be demonstrated![7]

At this point, one might think that the very idea of a *public* definition—something that articulates, in public language, what it is for something to be black or even publicly shows what it is—is problematic. The problem, one might think, is that the property of being black, as each of us grasp it, is essentially tied to a certain phenomenal quality that each of us is able to know, in our own case, but which we cannot describe with public language nor can we even publicly demonstrate, since we cannot know that it is instantiated by the experience that someone else has when they look at something that we communally call "black." So, each of us knows, considering our own experience, what *we* mean when we say that something is black, since we each know the quality of the experience *we* have when we see something black. It is *this quality*, understood in terms of its intrinsic phenomenal character, that we principally mean when we speak of "blackness." The property of being black, as a property of objects in the world, might be understood, in a secondary sense, as the propensity of objects to produce experiences that have this quality. Once again, there can be no *public* expression guaranteed to pick out this quality, for it may well be this is not the quality that your experiences instantiate when you look at objects that we both call "black." It could be, for instance, that, when you look at an object that we both call "black," your experiences instantiate the quality that my experiences instantiate when I look at an object we both call "white." Still, though there is no public expression that we can be sure to pick out this quality, we can nevertheless each coin a term for ourselves that directly picks out this quality in terms of its intrinsic character.[8] Thus, I might coin the term "X" to pick out the phenomenal quality instantiated by my experience when I see something we call "black," thereby providing the following "private definition" which specifies what *I*, at least, mean by "black":

$[\![\textbf{black}]\!]^{\text{R.S.}}$ = the property of being black (as I grasp it) =

The propensity of objects to produce experiences with quality X.

[7] This is what Wittgenstein (1953, §257) speaks of as the "stage setting" required for a successful ostensive demonstration.

[8] This view is most explicitly spelled out by David Chalmers (2010, 251-275).

I might likewise coin the term "Y" to pick out the phenomenal quality instantiated by my experience when I see something white:

$[\![\textbf{white}]\!]^{\text{R.S.}}$ = the property of being white (as I grasp it) =

The propensity of objects to produce experiences with quality Y.

I will leave it as an exercise for the reader to define such semantic values for themself and show, to their own satisfaction, that they cannot actually make determinate sense of what is purportedly expressed by their terms "X" and "Y," at least, not without appealing to their grasp of what is expressed by the public expression "black" and "white," thus bringing us right back to the problem with which we started.[9]

We have gone down quite a path in our attempt to say what the properties that figure into our simple intra-worldly semantic theory for our toy langauge actually are. By this point, one might have come to the conclusion that any attempt to provide a proper specification of what these properties are, at least for the basic ones like the property of being black, is bound to fail. But this would be too hasty. Let me now define these properties, thereby showing that they really can be defined.

3.6 The Way to Define Properties

There is, I think, a way that properties can be defined, though I take it that these definitions will always be relative to the rules of a linguistic practice with a particular structure. To see what I mean here, consider the properties grasped by the hypothetical speakers of our toy language. If we can imagine the speakers of our toy language as cognizers at all, we must suppose that there is some sense in which they grasp the property of being black, the property of being gray, the relation of being darker than, and so on. Furthermore, if we can imagine the speakers of our toy language as cognizers at all, then, for any property or relation that they grasp, there must be some specification of that property in the very terms which they themselves grasp it. Here is a proposal, based on this idea:

[9] If the reader would like some help with this exercise, I point them, first, to Wittgenstein's *Investigations* §28–§39,§239–§304, and, second, if help is still needed, to Stroud's (2002) very helpful guide, "Wittgenstein's 'Treatment' Of the Quest for 'A Language Which Describes My Inner Experiences and Which Only I Myself Can Understand.'"

⟦black⟧ = the property of being black =

> The property such that, if something instantiates it, then, necessarily, it is darker than anything gray or white, nothing is darker than it, everything is either the same shade as it or lighter than it, and so on.

Here, we've supplemented the vocabulary of the speakers of our toy language with some additional logical vocabulary: words like "if," "then," "necessarily," "anything," "nothing," and "everything."[10] With this additional logical vocabulary, the speakers of our toy language are able to specify not only the objects that instantiate the property of being black (though, importantly, they can also do that for at least some of them) but also the modal relations that this property stands in to other properties and relations expressible in their language. The proposal is that the property of being black just is what is expressed by the above sentence of the logically enriched toy language, namely, a bit of metaphysical structure. By "metaphysical structure" I mean nothing but that structure which can aptly be expressed with a metaphysical "necessity" operator, the sort of structure that, once we introduce the toolkit of possible worlds, we'll be able to articulate by universally quantifying over them. According to Sellars, this metaphysical structure is nothing but a codification of the exceptionless semantic norms governing the use of the predicate "is black."

In the next chapter, I'll give an official account of these "semantic norms," and, in the chapter after that, I'll give an official account of how logical vocabulary can function to make the semantic norms governing the use of predicates explicit. Here, however, I want to consider what the conception of the property of being black that is yielded by this final definition is, and why it is unavailable to our intra-worldly semanticist. For starters, on this definition, the property of being black is identified partly in terms of the modal relations that it bears to members of a family of related properties and relations. For instance, it is partly constitutive of what it is for something to be black on this definition that, for any objects x and y, if x is black and y is gray, then, necessarily, x is darker than y.

[10] Along with, among other things, the capacity for anaphoric reference. This is a bit of natural language that, though certainly essential for a full account of conceptual contents expressed by singular terms, we'll end up ignoring here for simplicity as our focus will be on properties, the conceptual contents expressed by predicates.

This is a modal relation that the property of being black stands in to the property of being gray and the relation of being darker than, and, on this definition, it is partly constitutive of what the property of being black is. Accordingly, if we opt for this definition, we can't appeal to what that property is—its "essence"—in order to explain the modal relations it bears to other properties and relations. But that, of course, is just what the intra-worldly semanticist proposes we do. Opting for this final definition constitutes a radical turn—the move from an *atomist* semantics for predicates, in which one explains the relations of entailment and incompatibility that obtain between the entities that are assigned to predicates as semantic values by appeal to independently intelligible features of these entities, to a *holist* semantics for predicates, in which the entities that are assigned to predicates as semantic values are only intelligible in virtue of the relations of entailment and incompatibility that they bear to one another.

To accept a holist semantics for predicates is a radical divergence from the sort of semantic theory we considered in the last chapter, in which the semantic relations obtain between predicates in virtue of these predicates' independently intelligible semantic values. Recall, on an extra-worldly semantic theory, the semantic values of "black" and "gray" are functions that map each possible world to a certain set of entities, the first function mapping each possible world to the set of black things in that world and the second function mapping each possible world to the set of gray things in that world. Simply given what these two functions are, it follows that, for each world, the set of entities to which the semantic value "gray" maps that world and the set of entities to which the semantic value of "black" maps that world are disjoint. Accordingly, given the definition of incompatibility provided by the extra-worldly semanticist, the extra-worldly semanticist can maintain that "black" and "gray" are incompatible in virtue of the specific semantic values of these two expressions. Now, of course, we raised a problem for semantic values of these sorts being advertised as models of properties, but we can nevertheless note that semantic values of these sorts do accord with a basic methodological principle: semantic relations that obtain between expressions of a language obtain in virtue of the semantic values of those expressions. Now, the intra-worldly semanticist recognizes the problem with semantic values of these sorts, seeing that our grasp of properties must be more

fundamental than the grasp of these functions. However, if, in attempting to define properties, the intra-worldly semanticist opts for this final definition, they cannot maintain this basic principle.

If one opts for this final definition, the meanings of predicates are understood, at least partly, in terms of their relations of implication and incompatibility that they bear to the meanings of other predicates. If one goes this route, it is a short step to the view that the properties that are taken to be the meanings of predicates are in part constituted by the semantic relations that those predicates bear towards other predicates. After all, it is clear from the failure of the third definition that the modal relations that we are permitted to appeal to in providing this final definition must be relative to the vocabulary of the speakers of the language for which we are providing a semantic theory. It is this specific class of modal relations that is partly constitutive of the properties that figure into the semantic theory. But what could this class of modal relations be other than the semantic relations that obtain between the predicates of the language? That, according this Sellarsian proposal, is just what properties are: codifications, in alethic modal terms, of semantic relations between predicates, where these semantic relations between predicates are just the relations of entailment and incompatibility that they stand to one another. This, I believe, is the correct theory of properties. The intra-worldly semanticist, however, cannot accept this theory, for, on their theory, properties are supposed to *explain* the relations of entailment and incompatibility that obtain between predicates of the language. So, we have here a familiar problem: if properties *explain* the relations of entailment and incompatibility that obtain between predicates, the relations of implication and incompatibility that obtain between predicates cannot *constitute* the properties. The intra-worldly semanticist must give a different account of the properties that figure in their semantic theory. But *what could that account be?*

At this point, it is worth recalling the basic dilemma faced by someone who has fallen prey to the Myth of the Given: they are stuck with a conception of our knowledge of some aspect of the structure of reality according to which it is either *unintelligible* or *incoherent*. It seems to me that the intra-worldly semantacist is stuck in just such a dilemma here. On the one hand, if they *don't* accept the account of properties I've just given, then, since the

account of properties I've just given is the only account that really can be given (because it is the correct one), then they have no account of properties. Accordingly, they're stuck with a semantic theory that is, at its base level, *unintellegible*. On the other hand, if they *do* accept the account of properties that I've given, then they appeal to the rules governing the use of predicates in order to account for the entities that are supposed to explain this use. Accordingly, they're stuck with a semantic theory that is, at it's core, *incoherent*. These are the only two options for the intra-worldly semanticist. Since, both options are unacceptable, so too is intra-worldly semantics.

3.7 Conclusion

Appeals to speakers' grasp of properties and relations is nearly universal in semantic theorizing. We saw, in the last chapter, that attempts to define such entities as constructions from possible worlds either make it impossible to understand how speakers grasp such entities, ending up with an account that's unintelligible, or appeal to speakers' grasp of properties in explaining this grasp, ending up with an account that's incoherent. In this chapter, we considered theories that don't define properties in terms of possible worlds, but, rather, take properties as basic, appealing to speakers' grasp of properties, primitively construed, to explain their knowledge of linguistic meaning. We have now argued that such theories face a similar problem: either one is left with no account of these properties at all or an account in which they are understood in terms of the very thing that they are supposed to help explain. Though we haven't considered all possible forms of worldly semantics, we have considered enough, I take it to license the main negative conclusion of this dissertation: worldly semantics is committed to the Myth of the Given. Given the constraints of these theories, the worldly knowledge to which such theories appeal cannot be understood except as *simply given* to speakers of a language, and that is not actually a way to understand knowledge at all. To avoid the Myth, we must turn to a very different sort of semantic theory, one that does not presuppose this sort of worldly knowledge but, rather, actually enables us to account for it. That is the positive task to which we now turn.

4

Discursive Role Semantics

4.1 Introduction

In the previous two chapters, I argued against worldly semantics in its two most prominent forms—what I called "extra-worldly" semantics and "intra-worldly" semantics. I claimed that worldly semantic theories of both sorts are not able to explain our knowledge of meaning because the worldly knowledge to which they appeal can only be understood as depending on our knowledge of meaning. I now turn to the positive task of explicating the sort of semantic theory that can do what worldly semantics cannot: explain our knowledge of meaning and, along with it, our knowledge of the "worldly" entities, such as possible worlds and properties, to which worldly semantic theories centrally appeal. This "worldly" knowledge, on the account I develop, is conceived of as nothing other than semantic knowledge, transposed into a worldly mode. The task of this chapter it to lay out the alternate, non-worldly semantic theory—based on the semantic theory proposed by Sellars (1953, 1954, 1974) and developed by Robert Brandom (1994)—that sets the ground for this account of "worldly" knowledge put forward in the next chapter. On this semantic theory, which I call "discursive role semantics," the meaning of an expression is understood directly in terms of its role in discourse, rather than this role being understood in terms of the sentence's having the worldly meaning that it does.

4.2 A Different Kind of Semantic Theory

Discursive role semantics is an alternative to truth-conditional semantics. As such, perhaps the best way of introducing it is to introduce it as a member of a wider class of

alternatives to truth-conditional semantic theories that have gained some traction in the past few decades: *dynamic* semantic theories. A dynamic semantic theory is one in which, rather thinking of the meaning of a sentence in terms of the conditions under which it is true, we think of the meaning of a sentence in terms of its potential, when employed in a given context, to change (or "update") that context. In a slogan, the meaning of a sentence is its context change potential. The meaning of a subsentential expression is the contribution that it makes to the context change potential of sentences in which it can occur.

On a standard dynamic theory, we think of the contexts that get updated when sentences are employed as information states.[1] The basic idea is that discourse participants are in certain information states at a given point in discourse, and the use of a particular sentence by some discourse participant will change the informational states of all the participants in that context who accept that sentence. If the sentence is informative, the participants will possess information that they had previously not possessed. In a simple sort of update semantics, we might model the information common to all parties as the set of worlds that are epistemically possible given their information (the set of worlds that, so far as these participants know, could be actual). We can then think about an update, potentially effected upon the assertoric utterance of some sentence φ in some context σ, as a mapping from the set of worlds that are taken to be epistemically possible by the participants of σ before the utterance of φ to the set of worlds that are taken to be epistemically possible after. Assuming that everyone in the context is in the same information state at a given point in discourse (an assumption we may eventually want to drop), we have a semantic theory in which the value assigned to φ is a function $[\varphi]$ that maps each discursive context σ in which φ can be employed to the context $\sigma[\varphi]$ that would result upon its being employed in that context.

A dynamic theory of this sort is a possible worlds semantics, but one in which the values of sentences are not sets of worlds, but, rather, functions from sets of worlds to sets of worlds. For an atomic sentence p, updating σ with $[p]$ results in a context $\sigma[p]$ that contains only the worlds in σ in which p is true. That is:

[1] See Veltman (1996), Groenendijk, Stokhof and Veltman (1996), Gillies (2004), and Willer (2013).

1. $\sigma[p] = \{w \in \sigma : p \text{ is true in } w\}$

The update effected by the assertoric use of a sentence of the form $\neg\varphi$ in context σ has the opposite effect, subtracting the $\sigma[\varphi]$, the context that would result from the assertoric use of φ in σ, from σ. That is,

2. $\sigma[\neg\varphi] = \sigma - \sigma[\varphi]$

Conjunction is treated as sequential update. So, the update effected by the assertoric use of a sentence of the form $\varphi \wedge \psi$ in context σ is one in which σ is first updated with $[\varphi]$ and then updated with $[\psi]$. That is,

3. $\sigma[\varphi \wedge \psi] = (\sigma[\varphi])[\psi]$

Disjunction can be defined in terms of negation and conjunction, exploiting the idea that we can think of $\varphi \vee \psi$ as $\neg(\neg\varphi \wedge \neg\psi)$. So, though it's a bit unwieldy, we have:

4. $\sigma[\varphi \vee \psi] = \sigma - ((\sigma - \sigma[\varphi]) - (\sigma - \sigma[\varphi])[\psi])$.[2]

Recursively defining updates of logically complex sentences in this way enables us to specify the update function that is the semantic value of any logically complex sentence.

A standard "informational" dynamic theory of this sort takes it for granted that atomic sentences of the language encode pieces of information about how the world is. The assertoric utterance of an atomic sentence rules out a particular set of worlds in a given context in virtue of the fact that such a sentence encodes such a piece of information, one that can be modeled as a set of possible worlds—the worlds that are consistent with this information about how the world is. Standard informational dynamic semantics is thus, despite its differences from a "static" possible worlds semantics, still a worldly

[2]The derrivation is as follows:

$$\begin{aligned}\sigma[\varphi \vee \psi] &= \sigma[\neg(\neg\varphi \wedge \neg\psi)] \\ &= \sigma - \sigma[(\neg\varphi) \wedge (\neg\psi)] \\ &= \sigma - ((\sigma[\neg\varphi])[\neg\psi]) \\ &= \sigma - ((\sigma - \sigma[\varphi])[\neg\psi]) \\ &= \sigma - ((\sigma - \sigma[\varphi]) - (\sigma - \sigma[\varphi])[\psi])\end{aligned}$$

More complex dyanmic frameworks, for instance, billateral frameworks that assign both positive and negative updates such as that proposed by Willer (2021) are able to provide more elegant characterizations of disjunction.

semantics. Speakers' knowledge of worldly states of affairs and their relations is taken as basic with respect to their knowledge of the relations that obtain between sentences of their language. Consider just the notion of incompatibility between sentences. On a standard informational dynamic semantics, two sentences φ and ψ are incompatible just in case updating any context σ with $[\varphi]$ and then $[\psi]$ results in the absurd context, consisting in the null set of words. That is, φ and ψ are incompatible just in case, for all contexts σ, $(\sigma[\varphi])[\psi] = \varnothing$. So, for instance, "*a* is white" is incompatible with "*a* is black" just in case if you update any context with "*a* is white" and then "*a* is black," you end up with the absurd context, consisting in the null set of worlds. This will be the case just in case there is no world in which both "*a* is white" is true and "*a* is black" is true. So, if we want a theory of this sort to account for speakers' knowledge of the semantic relation between these two sentences, we must, in thinking of the incompatibility between these sentences in these terms, take it that speakers antecedently have knowledge of the fact that the set of worlds in which *a* is white and the set of worlds in which *a* is black are disjoint. In Chapter Two, I argued that we cannot appeal to knowledge of this sort in accounting for speakers' knowledge of meaning. So, endorsing a dynamic semantics of this informational variety does not evade the arguments against worldly semantics put forward in the previous two chapters.

The version of dynamic semantics I'll propose here, which does not appeal to a notion of information at all and so is not subject to the criticisms of worldly semantics put forward in the previous chapters, will be a formalization of the semantic theory proposed by Brandom (1994).[3] Brandom's basic idea is to start with a notion of discourse, understood along a game-playing model, and to then give an account of the propositional content of a sentence that can be used in that discourse has by thinking of the use of that sentence as a certain type of "discursive move," the significance of which can be understood entirely in terms of the change in score that making of such a move would bring about, as this change

[3]Perhaps more accurately, it's formalization of the *normative pragmatic* theory, put forth by Brandom in the first half of *Making It Explicit*, that is supposed to ground the *inferentialist semantic* theory, put forth in the second half of the book. Essentially, I'm avoiding this two-stage order of explanation and doing semantics directly in terms of the pragmatics. This is how Nickel (2013), who I'm drawing from, formulates things, and that's a general theme of dynamic semantics: thinking of semantic values in terms of the sort of update usually relegated to the pragmatics.

in score is assessed from the players of the game. Characterizing this semantic theory as a dynamic theory, it is one in which, rather than thinking of contexts as sets of possible worlds, we think of what a context is in terms of the "score" that characterizes a particular stage in discourse, and we think of the meaning of a sentence in terms of its potential to change that score.[4] The resulting framework is what Bernhard Nickel (2013) calls a "normative" rather than "informational" dynamic semantics, where the contexts that get updated are understood not in terms of the informational content they contain (modeled by a set of worlds), but in terms of the normative statuses that have been assigned to the discursive participants.

4.3 The Game-Playing Model of Discursive Practice

Following Brandom (1994, 2000), the basic idea of discursive role semantics is that we can model discourse on what he calls "the game of giving and asking for reasons." Uttering a sentence is conceived of as making a move in the game. Like any game, the game of giving and asking for reasons has rules. The basic rule in the game is that you can only make a move if you're *entitled* to make it. An entitlement is a sort of move-making license, something that's acknowledged by the players of the game as making a move available for one to make. There are a few ways in which one can acquire entitlement to a move. One way is to be attributed entitlement by another player who takes you to have made a move as the exercise of an entitlement-conferring capacity. One class of such capacities are reliable observational capacities, or instance, the capacity to see that something is the case. Such a capacity is entitlement-conferring in the sense that, insofar as one is taken to exercise it, one will be taken to be entitled to the claim one comes to endorse upon that exercise. Another way to come to be entitled to a move is by *inheriting* this entitlement from some other player who has licitly made the move. One of the main functions of actually making a move (as opposed to merely being entitled to make it) is that, in making

[4] Lewis (1979) proposed thinking of various aspects of meaning in terms of this sort of scorekeeping, but this sort of scorekeeping was taken to be a supplement to possible worlds semantics, not a replacement of it. Before Lewis, however, Sellars proposed to think of linguistic meaning entirely in terms of this sort of scorekeeping model, and it is this Sellarsian idea that gets taken up in Brandom's *Making It Explicit*.

a move to which you are entitled, you pass the entitlement that you have to make that move on to others, who are then able to make that move themselves.

What makes the game a game is that players can *challenge* each other's moves, calling into question the entitlement they have to a move that they've made. To respond to a challenge, you must demonstrate your entitlement to the move that was called into question. One way of responding to a challenge to some move of yours is to demonstrate how the making of that move was made available by way of your exercise of an entitlement-conferring capacity. If you're able to do such a thing, you've done what you need to do in order to secure your entitlement to that move, successfully defending your move against that challenge. In a case in which you've made your move on the basis of another player's making it, you can respond to a challenge by deferring back to this other player. It is then this player who must respond to the challenge, once again, either by demonstrating how the move that they made was made available to them by the exercise of an entitlement-conferring capacity, or by deferring to the authority of another player. Somewhere along the line, some player must have exercised an entitlement-conferring capacity, or else no one in that chain of deference is entitled to the move. The authority that you have in making a move, entitling other players to make it, corresponds to the responsibility that you are able to bear in responding to challenges to your making of that move. This is why, if you continually fail to be able to respond to challenges to moves that you've made, failing to live up to the responsibility that you've undertaken in making the moves that you have, other players will stop taking you to have any authority. Eventually, your moves will no longer be taken to have the significance that moves generally do, functioning to entitle others to make them.

Given this general structure of the game of giving and asking for reasons, we can think of what it is that you do in making some move as undertaking a particular sort of *commitment*—a commitment to demonstrate your entitlement to that move if challenged, either by showing how the making of the move was the product of the execution of an entitlement-granting procedure, or by deferring to another player from whom you've inherited the entitlement to it.[5] To undertake a commitment of this sort, in making some

[5]John MacFarlane (2010, 91) claims that this way of construing what it is to be committed to a move is

move, is to take on the responsibility that underwrites the authority that one has, in making a move, to entitle others to make that move. Since one has this authority only if one takes on this responsibility, undertaking a commitment in making a move is necessary in order for move-making to serve its basic function—entitling others to make the move that was made.

The key idea that enables us to construct a semantic theory on the basis of this conception of discursive practice is that when one undertakes a commitment to some move, say p, one will generally not commit oneself to that move and only that move. Rather, commitment to that move will bring with it commitments to certain other moves. These other commitments that one takes on in committing oneself to p are the *committive consequences* of p. If a move q is a committive consequence of a move p, then a player who commits herself to p is not only committed to p, but also committed to q. So, this player is not only responsible for defending p against potential challenges, but also responsible for defending q against potential challenges. A second, directly related relation is that of *permissive consequence*. Roughly, if q is a permissive consequence of p, a player who is committed and entitled to p is (prima facie) entitled to q and so can appeal to her entitlement to p in response to a challenge to q. Generally, if q is a committive consequence of p, q will also be a permissive consequence of p, though the converse might not hold.[6] So, though, in making some move, one commits oneself to more than just that one move, if one is entitled to the move one makes, one will also be able to appeal to this entitlement in response to a challenge to one of these other moves. Finally, commitment to some moves will preclude entitlement to others. These are what I'll call the *preclusive consequences* of

circular. The worry is that, if we think of what it is to be committed to a move p in terms of a commitment to demonstrate entitlement to that commitment, then the account of commitment to p must appeal to the notion of being committed to p. However, given the way I'm construing it here, there is no circularity involved. McFarlane's charge of circularity hinges on the claim that what one must be committed to demo. What one is committed to doing, on this construal, is demonstrating one's entitlement to *make the move*, not demonstrating one's entitlement to *be committed to the move*. When one makes a move one does, of course, undertake a commitment to it, but undertaking that commitment just is committing oneself to demonstrating one's entitlement to make that move.

[6]For instance, consider how certain inductive lines of reasoning might entitle one to some move on the basis of some other move, but not commit one to it. For instance, observing a red sky at night might, on inductive grounds, entitle one to the claim that the next day's weather will be fair, but one is presumably not *committed* to this claim upon being committed to the claim that the sky is red.

p.[7] If q is a preclusive consequence of p, then, if some player is committed to p, they are precluded from being scored as entitled to q, insofar as they remain committed to p.

I've specified three consequence relations. We now have the raw materials to specify, in broad outline, how the semantic theory constructed on the basis of this conception of discursive practice will work. To model the players' attribution of normative statuses to one another, we'll say that each player has a "scorecard" wherein she keeps track of all the normative statuses that she's attributed to all of the other players of the game. We'll call a player's attitudes of taking some moves to be committive, permissive, and preclusive consequences of others her "scorekeeping principles." When we model the meaning of some sentence, we'll do it from the perspective of some player who has a certain set of scorekeeping principles. We define a subset of the total set of sets of normative assignments, which are the scorecards that each player might have, given their scorekeeping principles. We can then define the semantic value of a sentence φ, from the perspective of some scorekeeper m, as a function that takes any player n and any scorecard σ that m might have, and returns another scorecard, $\sigma[\checkmark_n\langle\varphi\rangle]$, which is the result of m's updating σ with n's making the move φ. These are the candidates for models of the meanings of sentences that I think Brandom's semantic theory gives us, and what we have, defining semantic values in this way, is a kind of dynamic semantics. We take the semantic value of a sentence φ to be its context change potential. However, unlike standard informational dynamic theories in which contexts are understood as sets of epistemically possible worlds, contexts are understood in terms of sets of normative assignments that conform to a given speaker's scorekeeping principles—scorecards that that speaker might have.

Before getting into the details of the proposal, we are now in a position to see how the framework that Nickel proposes to systematize Brandomian semantics is inadequate. Nickel tries to model Brandomian contexts as sets of sentences; a given context, he says, can be understood as the set of sentences to which everyone in that context is committed

[7] Brandom calls these moves the moves that are "incompatible" with p. I avoid this terminology here, first, because it covers up the sense in which incompatibility is understood, in the first instance, in terms of this kind of normative consequence relation which is quite analogous to the relations of committive and permissive consequence.

(2013, 340). While this does lead to a simple semantic framework, it is not at all adequate to enable us to model a Brandomian context. Let me point out just two key problems with it. The first problem with this way of modeling a Brandomian context is that it assumes that there is some single set of claims to which everyone in the discursive context is committed. It is crucial, however, that it need not be the case that everyone in the context is committed to the same things. Indeed, on the Brandomian picture, communication requires that particular participants, who are uniquely entitled to claims, are able to uniquely commit themselves to claims, bearing the responsibility for those commitments such that players are able to pass the buck back to them. Only by seeing how commitments and entitlements are such as to vary from player to player are we able to see how the game can function as a model for communication at all. The second problem with this way of thinking about a Brandomian context is that only one normative status is considered, and, as I am sketching the framework, we cannot do without at least a few. This will be clear when it comes to giving a semantics for logical operators. For instance, negation will be defined in terms of preclusive consequence, where this notion requires the interplay between distinct normative statuses, the idea that *commitment* to some claim can preclude *entitlement* to another. So, though Nickel has the right idea for how to think about Brandomian semantics, the actual framework he proposes is utterly inadequate to put the Brandom's basic resources to work. With that noted, let us now flesh out the formal details of a more adequate formal semantic framework.

4.4 The Basic Framework

To start, it will be helpful to introduce a set of special symbols to express the normative statuses that one might bear towards a move:

1. "✓" expresses the status of *having made* a move. The formula $\checkmark_n\langle\varphi\rangle$ says that player n has made move φ. When this formula shows up in some player m's scorecard, this means that m scores n as having actually made the move φ.

2. "⊕" expresses the status of *being committed* to a move. The formula $\oplus_n\langle\varphi\rangle$ says that n is committed to φ. When this formula shows up in some player m's scorecard,

this means that m scores things in such a way that n is *obligated* to make φ if they're called upon to do so in the context of an appropriate challenge.

3. "$\bigcirc$" expresses the status of *being entitled* to a move. The formula $\Diamond_n\langle\varphi\rangle$ says that n is entitled to φ. When this formula shows up in some player m's scorecard, this means that m scores things in such a way that n is *permitted* to make φ insofar as they recognize that it's a move that they're in a position to make.

4. "$\ominus$" expresses the status of *being precluded from being entitled* to a move. The formula $\ominus_n\langle\varphi\rangle$ says that n is precluded from being entitled to φ. When this formula shows up in some player m's scorecard, this means that m scores things in such a way that n is *precluded* from licitly making φ, given the other moves that they've made.

Given a language $\mathcal{L}$, the set of moves appealed to in providing the discursive role semantics for $\mathcal{L}$ will simply be the set of sentences of $\mathcal{L}$, and the "players" appealed to in providing the discursive role semantics will be the speakers of $\mathcal{L}$. Thus, following Kukla, Lance, and Retall (2009) we can define the following:

> **Field of Play:** A *field of play* is a triple consisting of
>
> 1. A non-empty set of players (PLAYER)
> 2. A non-empty set of moves (MOVE)
> 3. A non-empty set of normative statuses (STATUS)

For our toy language, PLAYER is the set $\{A, B, C\}$, our three discursive participants, our three players of the game, MOVE is the set of all the claims that can be made by any of our three players, either by employing one of the 27 atomic sentences of our language or employing one infinite of the logically complex sentences, and STATUS is the set of statuses just defined: $\{\checkmark, \oplus, \bigcirc, \ominus\}$. The basic way in which these three elements come together is in the form of a *normative assignment*, defined as follows:

> **Normative Assignments:** A *normative assignment* is any formula consisting of the specification of an $\boldsymbol{s} \in$ STATUS, a $\varphi \in$ MOVE, and an $n \in$ PLAYER that is written as $\boldsymbol{s}_n\langle\varphi\rangle$.

For instance, "A is committed to $\langle b$ is gray$\rangle$" is a normative assignment that we write as $\oplus_A\langle b$ is gray$\rangle$. Now, to speak about normative positions that one might occupy, such as "being committed to $\langle b$ is gray$\rangle$," in abstraction from anyone's actually occupying that position, we'll introduce what we'll call a "player place-holder," for which we'll use the greek letter α. With this, we can define two more things:

> **Normative Positions:** A *normative position* is any formula consisting in an $\boldsymbol{s} \in$ STATUS, a $\varphi \in$ MOVE, and a player place-holder, which is written as $\boldsymbol{s}_\alpha\langle\varphi\rangle$.
>
> **Scorekeeping Principles:** A *scorekeeping principle* is a sequent of the form $\Gamma \vdash A$, where Γ is a (possibly null) sequence of normative positions, and A is a single normative position.

It needs to be emphasized that the use of the player place-holder α in the specification of scorekeeping principles is not to be understood in terms of universal quantification over the elements of PLAYER on the part of the speaker who has that scorekeeping principle. It will turn out to be the case that, for any scorekeeping principle that a speaker has, expressible with the use of this place-holder, there will correspond to a practice describable with the use of universal quantification. However, it important to be clear that one's *having* a scorekeeping principle, keeping score *in accordance* with it, is distinct from one's *making* a corresponding quantificational claim, *explicitly acknowledging* a commitment to the practice of keeping score, the specification of which would require universal quantification. One will only be able to do such a thing insofar as one's language contains quantificational vocabulary, and the semantic theory works by enabling us to comprehend such vocabulary as functioning to make *explicit* what must already be *implicit* in the practice of keeping score. So, it is worth emphasizing that the "α" in a scorekeeping principle is to be understood, in the first instance, as the generic "one" not the universal "everyone."

We can now define two fundamental sorts of consequence relations—committive and preclusive—as two different sorts of scorekeeping principles. Where Γ is any sequence of normative positions and φ is any element of MOVE, a principle of *committive consequence* is a scorekeeping principle of the form $\Gamma \vdash \oplus_\alpha\langle\varphi\rangle$, and a principle of *preclusive consequence*

is a scorekeeping principle of the form $\Gamma \vdash \ominus_\alpha\langle\psi\rangle$.[8] We will work on the simplifying assumption, which is fine for our toy language but will likely need to be reconsidered for a genuine natural language, that the main work in determining updates is done solely by principles of committive and preclusive consequence, and entitlement just comes along for the ride, being attributed whenever one commits oneself to something to which one is not precluded from being entitled. We'll say that a *material scorekeeping principle* is a scorekeeping principle containing only atomic sentences in the move spot of the normative positions it contains. Material scorekeeping principles are what determine the semantic significances of the atomic sentences. Sample material scorekeeping principles from our toy language include the following:

$\oplus_\alpha\langle a \text{ is gray}\rangle \vdash \ominus_\alpha\langle a \text{ is white}\rangle$
$\oplus_\alpha\langle a \text{ is black}\rangle, \oplus_\alpha\langle b \text{ is gray}\rangle \vdash \oplus_\alpha\langle a \text{ is darker than } b\rangle$
$\oplus_\alpha\langle a \text{ is darker than } b\rangle, \oplus_\alpha\langle b \text{ is darker than } c\rangle \vdash \oplus_\alpha\langle a \text{ is darker than } c\rangle$

Clearly, if we tried to enumerate all of the scorekeeping principles for our toy language in this way, it'd be quite a long list! The number of scorekeeping principles we'll have will be reduced once we articulate the theory at a subsentential level. Once we do that, our principles will be general, not just with respect to the player expressions they contain, but with respect to the singular terms occurring in the specifications of the moves, and with that sort of generality we will then be able to easily specify all the principles we need for this simple toy language. We'll do that in Section 4.7. For the moment, however, we'll stay at the level of sentences to get an initial grip on how the semantic theory is meant to work.

Scorecards get updated through the application of scorekeeping principles like these. Consider just the first: $\oplus_\alpha\langle a \text{ is gray}\rangle \vdash \ominus_\alpha\langle a \text{ is white}\rangle$. The turnstile here can be informally understood as saying that if some player is scored as occupying the positions on the left, then they are to be scored as occupying the position on the right. So, applying this scorekeeping principle to a scorecard σ amounts to seeing if σ contains $\oplus_n\langle a \text{ is gray}\rangle$ for any player n, and, if it does, adding $\ominus_n\langle a \text{ is white}\rangle$ to σ, scoring anyone who one scores as

[8] Note that this a generalization of Brandom's own definitions as we allow both commitments and preclusions on the left of the turnstile.

committed to "a is gray" to be precluded from being entitled to "a is white." To officially state this idea, where A is some normative position of the form $\boldsymbol{s}_\alpha\langle\varphi\rangle$, let us use the notation A_n to denote the result of substituting the player place-holder α with a player n. Likewise, for a set of positions Γ, let Γ_n denote the result of substituting the player place-holder in each position in Γ with n. We can then define the application of principles as follows:

> **Application of Principles:** The result of applying a set of scorekeeping principles π to a scorecard σ, which we denote $\pi(\sigma)$, is the smallest superset of σ such that for every principle of the form $\Gamma \vdash A \in \pi$ and every player n, if $\Gamma_n \in \pi(\sigma)$, then $A_n \in \pi(\sigma)$

This definition of application of scorekeeping principles ensures that the operation of applying a set of scorekeeping principles to a scorecard is a closure operation. That is, for any scorecards σ and τ, the following facts hold:

> **Extensivity:** $\sigma \subseteq \pi(\sigma)$
> **Monotonicity:** If $\sigma \subseteq \tau$, then $\pi(\sigma) \subseteq \pi(\tau)$
> **Idempotency:** $\pi(\pi(\sigma)) = \pi(\sigma)$

Thus, a set of scorekeeping principles can be understood much like a classical consequence relation, under which a set of sentences, or in this case, normative assignments, can be closed.

We can now define two things, relative to one another: a set of scorecards that each player m might have, and the effect of any player n's making some move φ, relative to any scorecard that m might have:

> **Scorecards Players Might Have:** Let m be any player with a set of scorekeeping principles π. The set of scorecards Σ_m that m might have can be recursively defined as follows:
>
> 1. $\varnothing \in \Sigma_m$
> 2. For any $\sigma \in \Sigma_m$, any $n \in$ PLAYER, and any $\varphi \in$ MOVE, $\sigma[\checkmark_n\langle\varphi\rangle] \in \Sigma_m$
>
> **Updates:** Let n be any other player. The result of updating σ with $\checkmark_n\langle\varphi\rangle$, which we write as "$\sigma[\checkmark_n\langle\varphi\rangle]$," is defined as the final step in the following three step process:
>
> 1. $\sigma[\checkmark_n\langle\varphi\rangle]_1 = \sigma \cup \{\checkmark_n\langle\varphi\rangle, \oplus_n\langle\varphi\rangle\}$
> 2. $\sigma[\checkmark_n\langle\varphi\rangle]_2 = \pi(\sigma[\checkmark_n\langle\varphi\rangle]_1)$

3. $\sigma[\checkmark_n\langle\varphi\rangle] = \sigma[\checkmark_n\langle\varphi\rangle]_2 \cup \{\bigcirc_n\langle\psi\rangle\}$ for any $\psi \in$ MOVE such that $\oplus_n\langle\psi\rangle \in \sigma[\checkmark_n\langle\varphi\rangle]_2$, and neither $\ominus_n\langle\psi\rangle \in \sigma[\checkmark_n\langle\varphi\rangle]_2$ nor $\ominus_m\langle\psi\rangle \in \sigma[\checkmark_n\langle\varphi\rangle]_2$

So, supposing we are m, we assume that one way that we might score the game is to have it such that no one has played any moves at all, and so no one is committed, entitled, or precluded from being entitled to anything. When some player n makes some move φ, we add n's having made φ and being committed to φ to our scorecard. We then apply our scorekeeping principles to that scorecard, assigning to n any positions that follow from our scorekeeping principles. Finally, we attribute entitlement to any move ψ to which we now score as n as committed, unless we take n to be precluded from being entitled to ψ or we take ourselves to be precluded from being entitled to ψ. This last step amounts to Brandom's (1994, 176-178) principle of "default entitlement," according to which when one makes a claim one is generally taken to be entitled to it by default, unless there's some specific reason to challenge it, such as incompatibility with the claimant's commitments or our own, and that's how entitlement figures into the system here. So, the way we are doing things here, scorekeeping principles fundamentally involve the attributions of commitments and preclusions of entitlements, and entitlement just comes along for the ride by default wherever it can.

Defining updates and scorecards players might have in this way lets us define the semantic value of a sentence φ, relative to a player m, as a function that takes any scorecard m might have and any other player n and returns the scorecard that is the result of m's updating their scorecard with n's making move the φ:

Semantic Values:

$$[\![\varphi]\!]^m = f : (\Sigma_m \times \text{PLAYER}) \rightarrow \Sigma_m$$
$$f(\sigma, n) = \sigma[\checkmark_n\langle\varphi\rangle]$$

There will be an important difference between the semantic values of this sort that we'll define here, using the sentences of our toy language as an example, and the semantic values defined by the semantic frameworks previously considered. Those frameworks were *compositional* in the strong sense that the semantic values of complex expressions were built out of the semantic values of their parts. It is widely thought that this sort

of compositionality is necessary in order to explain the *productivity* of language, the fact that speakers, who clearly can have only a finite amount of knowledge, are capable of understanding a potentially infinite number of complex sentences. Strong compositionality, however, is not actually necessary to explain this fact, and it is not a feature of the semantic theory presented here. Rather than accounting for this fact by thinking of the meanings of these sentences as composed out of meanings of the parts, we account for this fact by thinking of the rules for determining the semantic significance of a sentence as *recursive*, such that rules for keeping score on the utterances of expressions of arbitrary complexity can be determined by the rules for keeping score on simple expressions. The recursive determination of meanings is all that's necessary to account for the fact that speakers can understand a potentially infinite number of sentences. We need not think of meanings themselves as compositional, in the sense of being composed out of the meanings of their parts.[9] Thus, our definition of semantic values will be *recursively* determined without being *compositionally* determined. The task of defining semantic values for the total set of sentences of the language amounts to the task of recursively specifying rules for keeping score, such that, given a base set of scorekeeping principles, which determine the semantic significance of the simple expressions of the language, one can specify a set of scorekeeping principles sufficient for determining the semantic significance of any of the complex expressions belonging to the language. Let us first consider how, given a set of scorekeeping principles that relate positions involving atomic sentences, we can determine the set of scorekeeping principles that relate positions involving any of the logically complex sentences.

4.5 Introducing Logical Operators

If we're giving a discursive role semantics for some language, we'll start by specifying our set of scorekeeping principles concerning only the atomic sentences of that language. This enables us to specify the update function that is the semantic value of each atomic sentences. In order to extend the semantics to logically complex sentences, we need a way

[9] See Brandom (2008, 133-136) for a discussion of this point.

of extending our set of scorekeeping principles so that we can specify the update function for logically complex sentences. Nickel (2013), who thinks of Brandomian contexts in terms of sets of sentences that everyone is committed to, thinks that specifying such updates will be quite difficult. He writes,

> Conjunction is easy: a speaker who asserts a conjunction $p \wedge q$ and thus commits herself to it just commits herself to each of the conjuncts p and q. Negation is trickier: committing oneself to $\neg p$ is not the same as not committing oneself to p—the latter, but not the former, is compatible with agnosticism about p, (345).

Nickel is right that, if the only status we have is commitment, defining negation in normative terms is tricky, indeed, probably impossible.[10] But if we have multiple normative statuses—particularly, the statuses of commitment and preclusion of entitlement—it is relatively straightforward.

Consider Brandom's (1994, 2008) definition of the negation of a sentence φ as its "minimal incompatible," the sentence implied by every set of sentences incompatible with φ. Now, on Brandom's definition of incompatibility in terms of scorekeeping, a set of sentences Γ is incompatible with φ just in case *commitment* to all the sentences in Γ precludes *entitlement* to φ. Bringing these two ideas together, we can introduce scorekeeping principles which attribute commitment to a negation by saying that, if occupying a set of normative positions Γ precludes one from being entitled to some sentence φ, then Γ commits one to its negation, $\neg\varphi$. That is:

$$\frac{\Gamma \vdash \ominus_\alpha \langle \varphi \rangle}{\Gamma \vdash \oplus_\alpha \langle \neg\varphi \rangle} \oplus_\neg$$

Alternately, if Γ commits one to φ, then Γ precludes one from being entitled to $\neg\varphi$:

[10] This is something that Nickel himself doesn't seem to realize. When he tries to consider Brandom's (2008) incompatibility semantics for logical operators, he isn't even able to define the notion of incompatibility on which Brandom's semantics is based. Nickel tells us, purporting to speak for Brandom, "two sentences are incompatible just in case commitment to one precludes commitment to the other," (345). This is crucially not Brandom's definition of incompatibility. For Brandom, two sentences are incompatibile just in case commitment to one precludes *entitlement* to the other. One surely *can* be *committed* to two incompatible sentences. What one *can't* be is *committed and entitled* to both sentences, since commitment to one precludes entitlement to the other. Taking there to be two normative statuses that interact in this way is one of the fundamental technical innovations of Brandom's semantic theory that distinguishes it over predecessor theories of a similar theoretical orientation, most notably Dummett's (1991) semantics based solely on the notion of entitlement (assertability), and this interplay between the statues of commitment and entitlement is absolutely essential to Brandom's definition of incompatibility and, accordingly, negation.

$$\frac{\Gamma \vdash \oplus_\alpha\langle\varphi\rangle}{\Gamma \vdash \ominus_\alpha\langle\neg\varphi\rangle}\ \ominus_\neg$$

These rules, which capture Brandom's conception of negation as the minimal incompatibility, with incompatibility being understood in scorekeeping terms, are just the introduction rules proposed in Ian Rumfitt's (2000) bilateral natural deduction system. The basic formal idea of bilateral logic, proposed by Rumfitt (2000) and Timothy Smiley (1996) before him, is that we associate each sentence of the language with a sign, positive or negative. Smiley and Rumfitt think of these two signs in terms of two opposite acts, a positive act of acceptance or affirmation and a negative act of rejection or denial. We'll provide a different interpretation of these signs, thinking of the "two ways" of bilateral logic in terms of two opposing normative statuses that one might have with respect to some move. Where φ is a move, there are two opposing non-neutral ways in which one might stand, normatively, with respect to it: one might be *committed* to it, or one might be *precluded from being entitled* to it.[11]

Now, there are different bilateral systems for the logical connectives that will fulfill our purpose, and any off-the-shelf bilateral system for classical logic, such as the natural deduction systems proposed by Smiley (1996) or Rumfitt (2000), will do.[12] However, since what we need is a way to *expand* a set of scorekeeping principles relating atomic sentences to a set of scorekeeping principles relating logically complex sentences, our purposes are really better fulfilled by a sequent calculus, along the lines of Gentzen's (1935/1969) classical sequent calculus LK, where we only have introduction rules. Now, Gentzen's LK has *multiple conusions*; whereas the premises of a sequent of the form $\Gamma \vdash \Delta$ are interpreted *conjunctively*, the conclusions are interpreted *disjunctively*. Rather than

[11] It is not hard to see why there is this correspondence between these normative statuses and the two signs of Rumfitt's logic. If one is committed to a move, then, when prompted to affirm or deny the move, one is committed to affirming it. If one is precluded from being entitled to a move, then, when prompted to affirm or deny the move, one is committed to denying it. I develop these ideas in more detail in Simonelli (M.S.b.).

[12] Rumfitt's system (2000, 800-802), is arguably more natural than Smiley's, but contains twice as many rules. It's worth pointing out in connection that one benefit of defining semantic values in the dynamic way that we have, rather than as they are defined in proof-theoretic semantics (Francez 2015, Stovall 2021), is that we won't have differences in meaning depending on which of two systems that both determine the same consequence relation we pick. On our approach, it is the consequence relation determined by the logical rules, which determines updates, that matters in defining semantic values, rather than the logical rules themselves.

having multiple conclusions like Gentzen's LK, our bilateral system will have only *single conclusion* sequents, and, rather than having rules for introducing connectives on the left and right of the turnstile, we'll have positive and negative rules: rules for attributing *commitments* and *preclusions of entitlements*.[13] This not only provides a more intuitive calculus, from the perspective of a traditional understanding of consequence, avoiding, for instance, Rumfitt's (2008) criticisms of multiple conclusion sequent calculi, but it is also technically crucial here, since we want to understand sequents in terms of their role in *updating* scorecards. Only a single conclusion sequent of the form $\Gamma \vdash A$ positively provides a scorekeeper with an instruction of *what to do* when they score someone as occupying all of the positions in Γ: score them as occupying the position A. If we want intuitive rules that define the classical connectives, and function to expand scorekeeping principles from atomic to logically complex sentences, a bilateral sequent calculus is just what we need.[14]

In addition to being bilateral and having single conclusions, the sequent calculus we'll provide will differ from Gentzen's in another crucial way: rather than having only *logical* axioms we'll also have *material* axioms—namely, any of the material scorekeeping principles.[15] So, where Γ is some set of normative positions, and A is a single normative position, we have the following axiom schema:

$$\frac{}{\Gamma \vdash A}\ \text{Material Base (MB)}$$

if $\Gamma \vdash A$ is a material scorekeeping principle

We'll also have an axiom that says, trivially, that if you score someone as occupying some set of normative positions, then you score them as occupying any normative position in that set. Call this axiom *Containment* (Brandom 2018):

[13]The idea of a bilateral sequent calculus of this sort is, as far as I'm aware, a new one. Bilateral logic has standardly been proposed in the form of natural deduction systems, which have both introduction and elimination rules (Smiley 1996, Rumfitt 2000, Francez 2015). Gentzen's multiple conclusion sequent calculus and related systems have notably been *interpreted* bilaterally (Restall 2006, Ripley 2013), with a sequent of the form $\Gamma \vdash \Delta$ interpreted as saying that affirming everything in Γ and denying everything in Δ is "out of bounds," but these sequent calculi have not themselves been tweaked so that the turnstile relates positively or negatively signed formulas, as has been done in the case of natural deduction systems.

[14]See Simonelli (M.S.b), once again, for a fuller development of these ideas.

[15]For a discussion of how Gentzen-style systems can be put to this use, see Brandom (2018), and for some examples, see Hlobil (2017) and Kaplan (2018). The specific version of this approach presented here was developed in collaboration with the ROLE (Research on Logical Expressivism) group, led by Brandom and Hlobil, as I discuss in the next chapter.

$$\frac{}{\Gamma, A \vdash A} \text{ Containment (CO)}$$

In both of these schemas, we require that Γ and $\{A\}$ contain only normative positions relating one to atomic sentences. Finally, since we're taking what goes on the left side of a sequent of the form $\Gamma \vdash A$ to be a *set* of normative positions, it doesn't matter how many times a normative position appears on the left of a sequent—the sequent expresses the same scorekeeping principle. So, we have the following two structural rules:

$$\frac{\Gamma, A, A \vdash B}{\Gamma, A \vdash B} \text{ Contraction (CNT)} \qquad \frac{\Gamma, A \vdash B}{\Gamma, A, A \vdash B} \text{ Expansion (EXP)}$$

Moreover, for the same reason, the order of normative positions on the left of a sequent doesn't matter:

$$\frac{\Gamma, A, B, \Delta \vdash C}{\Gamma, B, A, \Delta \vdash C} \text{ Permutation (P)}$$

One can confirm that our definition of the application of scorekeeping principles validates these structural rules in the sense that closing a set of scorekeeping principles under any of these structural rules does not have any effect on the update effected by applying that set of scorekeeping principles to a scorecard.

In addition to these structural rules, one crucial bilateral structural rule is necessary for the system I'll lay out to work. Note that the conception of negation here is based on the notion of *incompatibility*, understood in terms of the normative statuses commitment and preclusion of entitlement, and it's crucial to a proper understanding of this notion, as well as these connective rules we'll give, that incompatibility is *symmetric*.[16] This is clearly the case if we consider some concrete examples. For instance, commitment to "*a* is gray" precludes entitlement to "*a* is white," and, just as well, commitment to "*a* is white" precludes entitlement to "*a* is gray." Generally, if $\Gamma, \oplus_\alpha\langle\varphi\rangle \vdash \ominus_\alpha\langle\psi\rangle$, then $\Gamma, \oplus_\alpha\langle\psi\rangle \vdash \ominus_\alpha\langle\varphi\rangle$. This is a sort of bilateral contraposition principle, and we can note that the goodness of this sort of contraposition principle generalizes. For instance, not only is the relation *contrariety*

[16]The symmetry of incompatibility is presupposed by Brandom's (1994) definition of incompatibility. It is also explicitly assumed in incompatibility-based semantics for non-classical logics proposed by Restall (1999) and Berto (2015), though this assumption has been questioned by De and Omori (2018). I've argued elsewhere (Simonelli M.S.c) that we need not take this fact as simply primitive; we can actually give a pragmatic argument for why incompatibility *must* be symmetric.

symmetric, but so is the relation of *subcontraiety*, where φ and ψ are subcontraries, relative to a set of positions Γ, just is case being precluded from being entitled to φ commits one to ψ, and vice versa. Consider, for instance, that, relative to commitment to "*a* is a primary color" and preclusion of entitlement to "*a* is blue," preclusion of entitlement to "*a* is yellow" commits one to "*a* is red," and, just as well, relative to the same set of positions, preclusion of entitlement to "*a* is red" commits one to "*a* is yellow." So, generally, if $\Gamma, \ominus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle$, then $\Gamma, \ominus_\alpha\langle\psi\rangle \vdash \oplus_\alpha\langle\varphi\rangle$. Generalizing this sort of contraposition, we get the structural rule that Smiley (1996) dubs *Reversal*. Where A and B are normative positions and starring a normative position yields the opposite signed normative position (such that, if A is of the form $\oplus_\alpha\langle\varphi\rangle$, then A^* is $\ominus_\alpha\langle\varphi\rangle$ if vice versa), the rule can be stated as follows:

$$\frac{\Gamma, A \vdash B}{\Gamma, B^* \vdash A^*}\ \text{Reversal}$$

The simple structural rules which all follow directly from our scorekeeping interpretation of the sequents, along with with this bilateral structural rule, are all the structural rules we need for this system.[17]

Let us now state the rest of the connective rules, the positive and negative conjunction and disunction rules. Consider first the positive conjunction rule. If a set of normative positions Γ commits one to φ, and Γ also commits one to ψ, then Γ commits one to $\varphi \wedge \psi$:

$$\frac{\Gamma \vdash \oplus_\alpha\langle\varphi\rangle \quad \Gamma \vdash \oplus_\alpha\langle\psi\rangle}{\Gamma \vdash \oplus_\alpha\langle\varphi \wedge \psi\rangle}\ \oplus_\wedge$$

Dually, for the negative disjunction rule, if a set of normative positions Γ precludes one from being entitled to ψ, and Γ also precludes one from being entitled to ψ, then Γ precludes one from being entitled to $\varphi \vee \psi$:

$$\frac{\Gamma \vdash \ominus_\alpha\langle\varphi\rangle \quad \Gamma \vdash \ominus_\alpha\langle\psi\rangle}{\Gamma \vdash \ominus_\alpha\langle\varphi \vee \psi\rangle}\ \ominus_\vee$$

Once again, these are just the introduction rules of Rumfitt's natural deduction system. The negative conjunction and positive disjucntion rules, however, are novel to this system.

[17] In fact, as I make clear in the Appendix, we don't even need Contraction and Expansion.

The negative conjunction rule says that if, relative to a set of normative positions Γ, φ and ψ are *contraries*, in the sense that commitment to one precludes entitlement to other, then Γ precludes one from being entitled to $\varphi \wedge \psi$:

$$\frac{\Gamma, \oplus_\alpha\langle\varphi\rangle \vdash \ominus_\alpha\langle\psi\rangle}{\Gamma \vdash \ominus_\alpha\langle\varphi \wedge \psi\rangle}\ \ominus_\wedge$$

Dually, if, relative to Γ, φ and ψ are *subcontraries*, in the sense that being precluded from being entitled to one commits one to the other, then Γ commits one to $\varphi \vee \psi$.

$$\frac{\Gamma, \ominus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle}{\Gamma \vdash \oplus_\alpha\langle\varphi \vee \psi\rangle}\ \oplus_\vee$$

The preclusive conjunction and committive disjunction are the new rules to this calculus, and they are not only essential to its technical workings, but also conceptually significant, in offering a new way of thinking about conjunction and disjunction in terms of contrariety and subcontrariety.

The sequent calclus constituted by these structural and operational rules, leaving out the material axioms, is a sound and complete system of classical logic. Any classical consequence will have a proof tree whose leaves are all instances of CO. Indeed, it's equivalent of Ketonen's (1944) formulation of Gentzen's LK, as I show in the Appendix, which is a very nice logic, from a technical perspective.[18] More importantly for our purposes, with these rules, a speaker's basic set of scorekeeping principles, involving only logically simple moves, can be expanded to include principles of committive and preclusive consequence with respect to logically complex claims. The basic idea underlying this way of introducing logical vocabulary is that to grasp this bit of vocabulary—to understand conjunction, disjunction, and negation—is to grasp how making a move in which it is used situates a player in the game. Accordingly, we can model the meaning of this bit of vocabulary by way of a set of rules which enable a player to expand their scorekeeping principles such that we can specify the update that takes place when a move in which

[18]The Ketonen system of which this is a translation has the same rules as the system Negri and von Plato (2008) call "G3cp," but with the standard negation rules of LK. See Curry (1963, 192-225) for a discussion of the system and some of its properties. The proof of the equivalence of Ketonen's system and the bilateral calculus presented here is provided in the Appendix.

this vocabulary is used is made. To see how these rules work, let's consider an example. We should want our rules for logical operators to combine with our basic scorekeeping principles in such a way that, since commitment to "a is black" precludes entitlement to "a is white," and commitment to "a is gray" precludes entitlement to "a is white," we'll have that commitment to "a is black or a is gray" commits one to "a is not white." We get this as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\oplus_\alpha\langle b\rangle \vdash \ominus_\alpha\langle w\rangle}{\oplus_\alpha\langle w\rangle \vdash \ominus_\alpha\langle b\rangle}\,\text{RV} \quad \dfrac{\oplus_\alpha\langle g\rangle \vdash \ominus_\alpha\langle w\rangle}{\oplus_\alpha\langle w\rangle \vdash \ominus_\alpha\langle g\rangle}\,\text{RV}}{\oplus_\alpha\langle w\rangle \vdash \ominus_\alpha\langle b \vee g\rangle}\,\ominus_\vee}{\oplus_\alpha\langle b \vee g\rangle \vdash \ominus_\alpha\langle w\rangle}\,\text{RV}}{\oplus_\alpha\langle b \vee g\rangle \vdash \oplus_\alpha\langle \neg w\rangle}\,\oplus_\neg$$

In this way, scorekeeping principles relating logically complex sentences can be generated by rules that expand a set of scorekeeping principles relating atomic sentences. In this way, we can make sense of one's ability to grasp the updates imposed by a potentially infinite number of complex sentences on the basis of a finite amount of knowledge—knowledge of the scorekeeping principles relating atomic sentences and knowledge of the rules for generating scorekeeping principles relating logically complex sentences.

4.6 Predicative Structure

In our very simple toy language, there is a basic syntactic distinction between two types of subsentential expressions: singular terms and predicates. This, of course, corresponds more closely to the syntactic structure of a simple formal language, like first-order logic, rather than a natural language like English, but the basic strategy of accounting for subsentential structure here can be extended to other syntactic categories.

Let us start with singular terms. The semantic significance of singular terms can be understood in terms of the way in which co-referential terms can be substituted in for one another with the discursive roles of the sentences they are substituted into being preserved. If a and b are taken to be co-referential by some scorekeeper m, then, for any scorekeeping principle m has in which a figures in a certain spot, either in the premises or

the conclusions, m will have a corresponding scorekeeping in which b figures in that spot. That is, m's scorekeeping principles will be closed under the following rules:[19]

$$\frac{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi(a)\rangle \vdash A}{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi(b)\rangle \vdash A} \qquad \frac{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi(b)\rangle \vdash A}{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi(a)\rangle \vdash A}$$

Of course, this yields a very simple account of the meaning of proper names, a Millian one that does not take into account anything like Fregean sense.[20] That is, indeed, something that this framework can accommodate due to its multi-perspectival nature. Following the account proposed in Chapter Eight of *Making It Explicit*, we can understand the differing senses of different co-referential singular terms (such as "Superman" and "Clarke Kent"), which may vary from perspective to perspective (for instance, from the perspective of Lois Lane to the perspective of Martha Kent) in terms of the different scorekeeping principles involving the substitution of these terms that different speakers have. However, rather than substantially complicating the formal framework to introduce such a system, since our main concern here is with the meanings of predicates, this simple account will do for our purposes.

Thinking about rules for substitution of this sort enables us to think abstractly about the roles of predicates in abstraction from any particular singular terms to which those predicates are attached. So, for instance, commitment to "a is gray" precludes one from being entitled to "a is white," and, if one is also committed to "b is black," commits one to "b is darker than a," and so on. To arrive at the roles of predicates, we consider that normative relations between these sentences stay constant if different singular terms are uniformly substituted with the singular terms they contain. So, we notice that, if we take another singular term, say "c," and substitute it for the utterance of "a" in any of these utterances, the normative relations between the moves made by the utterances are preserved. Thus, we can characterize the sentences "a is black" and "c is black," as both sentences of the form "x is black," and we can say, for instance, commitment to a sentence of the form "x is black," precludes one from being entitled to an sentence of the

[19] Because we have Reversal, these rules also give us the rules where the relevant formulas occur in the conclusion.

[20] This basic strategy of accounting for the inferential significance of proper names follows the proposal of Tanter (2021).

form "x is white," and, if one is additionally committed to a sentence of the form "y is gray," then one is committed to "x is darker than y," and so on. Talk of scorekeeping principles involving these "sentence forms" or, as Brandom puts it, *sentence frames* is intelligible through considering how the predicative aspect of sentential scorekeeping principles remains stable as different singular terms are substituted for one another into those principles.

So, finally, we can think of atomic scorekeeping principles as derived from scorekeeping principles relating sentence frames and rules for saturating those frames with singular terms. For instance, we can think of the scorekeeping principle on sentences $\oplus_\alpha\langle a \text{ is black}\rangle, \oplus_\alpha\langle b \text{ is gray}\rangle \vdash \oplus_\alpha\langle a \text{ is darker than } b\rangle$ as resulting from saturating scorekeeping principles on sentence frames $\oplus_\alpha\langle x \text{ is black}\rangle, \oplus_\alpha\langle y \text{ is gray}\rangle \vdash \oplus_\alpha\langle x \text{ is darker than } y\rangle$ with singular terms. To spell this out, let us suppose we can use a set of variables $x_1, x_2 \ldots x_n$, and we can think of any variable x_i as replaceable with a singular term by way of the following rule, where $\Phi_1, \Phi_2 \ldots \Phi_n$ and Ψ are any predicative contexts (which may or may not contain variables) and τ is any singular term belonging to the language:

$$\frac{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi_1(x_i)\rangle \ldots {}^{\oplus}/_{\ominus_\alpha}\langle\Phi_n(x_i)\rangle \vdash {}^{\oplus}/_{\ominus_\alpha}\langle\Psi(x_i)\rangle}{\Gamma, {}^{\oplus}/_{\ominus_\alpha}\langle\Phi_1(\tau)\rangle \ldots {}^{\oplus}/_{\ominus_\alpha}\langle\Phi_n(\tau)\rangle \vdash {}^{\oplus}/_{\ominus_\alpha}\langle\Psi(\tau)\rangle}$$

Thus, we have, for instance, the following double application of this rule:

$$\dfrac{\dfrac{\oplus_\alpha\langle x_1 \text{ is black}\rangle, \oplus_\alpha\langle x_2 \text{ is gray}\rangle \vdash \oplus_\alpha\langle x_1 \text{ is darker than } x_2\rangle}{\oplus_\alpha\langle a \text{ is black}\rangle, \oplus_\alpha\langle x_2 \text{ is gray}\rangle \vdash \oplus_\alpha\langle a \text{ is darker than } x_2\rangle}}{\oplus_\alpha\langle a \text{ is black}\rangle, \oplus_\alpha\langle b \text{ is gray}\rangle \vdash \oplus_\alpha\langle a \text{ is darker than } b\rangle}$$

In this way, we can think of the semantic significance of predicates in terms of the rules governing sentence frames that can be saturated with any singular terms, and this explains how, for instance, when a novel singular term, for instance "d," is used, a speaker will grasp the significance of saying "d is black," even though this is not a sentence they would have previously considered.

4.7 Providing the Full Lexical Semantics

With this formal machinery on the table, we can finally provide the full lexical semantics for our very simple toy language. To make this task a bit easier on ourselves, let us add the Structural rules of Monotonicity and Transitivity:[21]

$$\frac{\Gamma \vdash A}{\Gamma, B \vdash A} \text{ Monotonicity (MO)} \qquad \frac{\Gamma \vdash A \quad A \vdash B}{\Gamma \vdash B} \text{ Transitivity (T)}$$

Though these rules are not required for our logical system to work, one can confirm that these rules are also validated by our definition of updates, which ensures that updating is a closure operation. Now, if we want to formulate a version of discursive role semantics that would be adequate for a natural language like English, there will be reason to modify these definitions to go *substructural* so that we can have, for instance, $\oplus_\alpha\langle\textbf{bird}\rangle \vdash \oplus_\alpha\langle\textbf{flies}\rangle$ without having $\oplus_\alpha\langle\textbf{bird}\rangle, \oplus_\alpha\langle\textbf{penguin}\rangle \vdash \oplus_\alpha\langle\textbf{flies}\rangle$. For a substructural development of this sort of framework, see Simonelli (M.S.b, M.S.d). For our purposes here, however, things are simplified by treating things in this way.[22] In addition to these rules, let us add one more bilateral structural rule, which I'll call *Bilateral Reductio* (BR).[23] Once again, where A and B are any normative positions, and starring a normative position yields the oppositely signed position, the rule can be put as follows:

$$\frac{\Gamma, A \vdash B \quad \Gamma, A \vdash B^*}{\Gamma \vdash A^*} \text{ BR}$$

The idea is that if being committed to φ would leave one a situation in which one is both committed and precluded from being entitled to some sentence ψ, then one is precluded

[21] We can distinguish this Transitivity principle, which we might more carefully call "Simple Transitivity," from the weaker principle of *Cumulative* Transitivity:

[22] Even if we do go substructrual, however, we may nevertheless *locally* treat things this way, since, for this particular bit of vocabulary, the various structural rules *do* apply. See Hlobil (2017) for a discussion of this notion of structural rules holding locally.

[23] Rumfitt (2000) calls it *Smileian Reductio* (855) in reference to Smiley (1996) who first proposes the rule. It is perhaps worth noting that, given the translation procedure for going from this bilateral sequent calculus to a multiple conclusion sequent calculus (specified in the Appendix), that this principle corresponds directly to Gentzen's *Cut* rule:

$$\frac{\Gamma, \varphi \vdash \Delta \quad \Gamma \vdash \varphi, \Delta}{\Gamma \vdash \Delta} \text{ Cut}$$

from being entitled to φ. Likewise, if being precluded from being entitled to φ would leave one in such a situation, then one is committed to φ. Given BR, one can treat the Reversal rule specified above as a derived structural rule, derived as follows:

$$\dfrac{\dfrac{\dfrac{\Gamma, A \vdash B}{\Gamma, A, B^* \vdash B}\,\text{MO}}{\Gamma, B^*, A \vdash B}\,\text{P} \qquad \Gamma, B^*, A \vdash B^*\,\text{CO}}{\Gamma, B^* \vdash A^*}\,\text{BR}$$

Here too, if we go substructural, there will be reason not to include BR, just having RV as our bilateral structural rule, but, once again, things are simplified by treating them this manner.

We can now articulate a "kernel" from which we can derive the full lexical semantics for our toy language, articulating sixteen scorekeeping principles that a speaker of this language has from which all others can be derived:

1. $\oplus_\alpha\langle x \text{ is darker than } y\rangle \vdash \oplus_\alpha\langle y \text{ is lighter than } x\rangle$
2. $\oplus_\alpha\langle x \text{ is lighter than } y\rangle \vdash \oplus_\alpha\langle y \text{ is darker than } x\rangle$
3. $\oplus_\alpha\langle x \text{ is darker than } y\rangle, \oplus_\alpha\langle y \text{ is darker than } z\rangle \vdash \oplus_\alpha\langle x \text{ is darker than } z\rangle$
4. $\vdash \ominus_\alpha\langle x \text{ is darker than } x\rangle$
5. $\oplus_\alpha\langle x \text{ is the same shade as } y\rangle, \oplus_\alpha\langle y \text{ is the same shade as } z\rangle \vdash \oplus_\alpha\langle x \text{ is the same shade as } z\rangle$
6. $\oplus_\alpha\langle x \text{ is the same shade as } y\rangle \vdash \oplus\langle y \text{ is the same shade as } x\rangle$
7. $\vdash \oplus_\alpha\langle x \text{ is the same shade as } x\rangle$
8. $\oplus_\alpha\langle x \text{ is the same shade as } y\rangle \vdash \ominus\langle x \text{ is darker than } y\rangle$
9. $\oplus_\alpha\langle x \text{ is gray}\rangle, \oplus_\alpha\langle y \text{ is white}\rangle \vdash \oplus_\alpha\langle x \text{ is darker than } y\rangle$
10. $\oplus_\alpha\langle x \text{ is black}\rangle, \oplus_\alpha\langle y \text{ is gray}\rangle \vdash \oplus_\alpha\langle x \text{ is darker than } y\rangle$
11. $\oplus_\alpha\langle x \text{ is black}\rangle, \oplus_\alpha\langle y \text{ is white}\rangle \vdash \oplus_\alpha\langle x \text{ is darker than } y\rangle$
12. $\oplus_\alpha\langle x \text{ is white}\rangle, \oplus_\alpha\langle y \text{ is white}\rangle \vdash \oplus_\alpha\langle x \text{ is the same shade as } y\rangle$
13. $\oplus_\alpha\langle x \text{ is gray}\rangle, \oplus_\alpha\langle y \text{ is gray}\rangle \vdash \oplus_\alpha\langle x \text{ is the same sahde as } y\rangle$
14. $\oplus_\alpha\langle x \text{ is black}\rangle, \oplus_\alpha\langle y \text{ is black}\rangle \vdash \oplus_\alpha\langle x \text{ is the same shade as } y\rangle$
15. $\oplus_\alpha\langle x \text{ is black}\rangle \vdash \ominus_\alpha\langle y \text{ is darker than } x\rangle$
16. $\oplus_\alpha\langle x \text{ is white}\rangle \vdash \ominus_\alpha\langle y \text{ is lighter than } x\rangle$

Using our structural rules, we can derive the myriad other atomic scorekeeping principles from this basic set of scorekeeping principles. For instance, we can derive the principle

$\oplus_\alpha\langle x \text{ is gray}\rangle \vdash \ominus_\alpha\langle x \text{ is white}\rangle$

from (11) and (4) as follows:

$$\dfrac{\oplus_\alpha\langle Gx\rangle, \oplus_\alpha\langle Wx\rangle \vdash \oplus_\alpha\langle Dxx\rangle \quad \dfrac{\vdash \ominus_\alpha\langle Dxx\rangle}{\oplus_\alpha\langle Gx\rangle, \oplus_\alpha\langle Wx\rangle \vdash \ominus_\alpha\langle Dxx\rangle}\,\text{MO}}{\oplus_\alpha\langle Gx\rangle \vdash \ominus_\alpha\langle Wx\rangle}\,\text{BR}$$

and we can then derive the principle

$\oplus_\alpha\langle a \text{ is gray}\rangle \vdash \ominus_\alpha\langle a \text{ is white}\rangle$

as a particular instance of this general one, the instance in which the singular term "a" has been substituted into the open spot marked by "x." Similarly, we can derive:

$\oplus_\alpha\langle x \text{ is darker than } y\rangle \vdash \ominus_\alpha\langle x \text{ is lighter than } y\rangle$

from (2), (3), and (4) as follows:

$$\dfrac{\dfrac{\oplus\langle Dxy\rangle, \oplus\langle Dyx\rangle \vdash \oplus\langle Dxx\rangle \quad \dfrac{\vdash \ominus\langle Dxx\rangle}{\oplus\langle Dxy\rangle, \oplus\langle Dyx\rangle \vdash \ominus\langle Dxx\rangle}\,\text{MO}}{\oplus\langle Dxy\rangle \vdash \ominus\langle Dyx\rangle}\,\text{BR} \quad \dfrac{\oplus\langle Lxy\rangle \vdash \oplus\langle Dyx\rangle}{\ominus\langle Dyx\rangle \vdash \ominus\langle Lxy\rangle}\,\text{RV}}{\oplus\langle Dxy\rangle \vdash \ominus\langle Lxy\rangle}\,\text{T}$$

And, though it's a bit tedious, we can derive

$\oplus_\alpha\langle x \text{ is lighter than } y\rangle, \oplus_\alpha\langle y \text{ is lighter than } z\rangle \vdash \oplus_\alpha\langle x \text{ is lighter than } z\rangle$

from (1), (2), and (3) as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\oplus\langle Dzy\rangle, \oplus\langle Dyx\rangle \vdash \oplus\langle Dzx\rangle \quad \oplus\langle Dzx\rangle \vdash \oplus\langle Lxz\rangle}{\oplus\langle Dzy\rangle, \oplus\langle Dyx\rangle \vdash \oplus\langle Lxz\rangle}\,\text{T}}{\oplus\langle Dzy\rangle, \ominus\langle Lxz\rangle \vdash \ominus\langle Dyx\rangle}\,\text{RV} \quad \dfrac{\oplus\langle Lxy\rangle \vdash \oplus\langle Dyx\rangle}{\ominus\langle Dyx\rangle \vdash \ominus\langle Lxy\rangle}\,\text{RV}}{\oplus\langle Dzy\rangle, \ominus\langle Lxz\rangle \vdash \ominus\langle Lxy\rangle}\,\text{T} \quad \dfrac{\oplus\langle Lyz\rangle \vdash \oplus\langle Dzy\rangle}{\ominus\langle Dzy\rangle \vdash \ominus\langle Lyz\rangle}\,\text{RV}}{\oplus\langle Lxy\rangle, \ominus\langle Lxz\rangle \vdash \ominus\langle Lyz\rangle}\,\text{T}}{\oplus\langle Lxy\rangle, \oplus\langle Lyz\rangle \vdash \oplus\langle Lzx\rangle}\,\text{RV}$$

There is likely a simpler axiomatization of our toy language. Perhaps some of the scorekeeping principles included in this kernel can be derived from others in it, or perhaps there is a smaller set of material scorekeeping principles from which all of those included here can be derived. The point here is not to provide the simplest kernel, but simply to show that there is some not only finite but relatively manageable set of material scorekeeping principles from which the total set of material scorekeeping principles for this toy language can be derived.

Now, it is not clear whether we can actually provide anything like a full lexical semantics for natural language as we did with respect to our toy language. Nevertheless, several semantics, perhaps most notably Barbara Partee (2005), have suggested that the project of lexical semantics can be undertaken by doing something of this sort, laying down "meaning postulates." Such postulates, often formulated in first-order logic, can be interpreted straightforwardly in discursive role semantics as expressing basic scorekeeping principles. For instance, consider again the postulate which we first considered in Section 2.4:

$$\forall x(\mathbf{gray}(x) \rightarrow \neg\mathbf{white}(x))$$

On a model-theoretic way of thinking about meaning postulates, we might think of this postulate as saying, informally, that everything in the domain of discourse is such that if the predicate "gray" is correctly applied to it, it is not the case that the predicate "white" is correctly applied to it. In model-theoretic semantics, we might take these postulates, so interpreted, to constrain the models that are considered for the purpose of semantic theorizing. In the context of discursive role semantics, however, we can interpret it as expressing the following scorekeeping principle:

$$\oplus_\alpha\langle x \text{ is gray}\rangle \vdash \ominus_\alpha\langle x \text{ is white}\rangle$$

The idea of such a quantificational formula expressing such a material scorekeeping principle will be made precise in the next chapter. The point for now is just that a lexical semantics consisting in a set of meaning postulates for the atomic sentences of a natural language can be interpreted in this way, or, better, could be directly done in

this framework in terms of the explicit laying down of material scorekeeping principles. Discursive role semantics thus promises to provide a unified framework for both the lexical semantics of atomic sentences, understood in terms of a set of basic scorekeeping principles (which enable us to specify updates for the atomic sentences), and the proof-theoretic semantics for non-atomic sentences, understood in terms of rules for deriving scorekeeping principles (which enable us to specify updates for the non-atomic sentences), all within an overarching dynamic conception of meaning.

4.8 Conclusion

We have now presented a complete formal semantic theory for simple our toy language, understanding meaning in terms of discursive role. Though doing this has been a more substantial task than providing an extra-worldly or intra-worldly semantics for our toy language, what is important about the semantic theory we have laid out is that it, in principle, presupposes no worldly knowledge. I am actually prepared to make this claim unrestrictedly about worldly knowledge as such, but, for the purposes of the present project, the relevant sort of worldly knowledge that this semantic theory does not presuppose is knowledge of such things as properties, relations, and modal relations among them, or such things as sets of possible worlds and set-theoretic relations among them. Rather than the worldly contents expressed by predicates or sentences determining the relations of entailment and incompatibility that these predicates or sentences stand to one another, the relations of entailment and incompatibility between sentences are understood directly in pragmatic terms, in terms of commitment to some sentence committing one to others or precluding one from being entitled to others. This opens up the door for thinking about the "worldly" knowledge appealed to in the semantic theories previously considered—knowledge of modal relations between properties or set-theoretic relations between sets of worlds—as nothing other than knowledge of the norms governing the use of various expressions, but *reified*, transposed into a "worldly" mode. Spelling out such a conception of this worldly knowledge is the task to which I now turn.

5

"Worldly" Knowledge as Semantic Knowledge

5.1 Introduction

In the previous chapter, I showed how we can think of the meaning of a sentence in terms of what the utterance of that sentence *does*, normatively speaking, in a discursive practice in which it might be uttered. This enabled us to define semantic values of sentences, relative to the perspective of each speaker, as functions that map each scorecard that this speaker might have to the scorecard that would result upon some other speaker's uttering that sentence. These updates are determined by the various "scorekeeping principles" possessed by the speaker relative to which the updates are defined. The aim of this chapter is to show how we can think of modalized quantified conditionals, like "If something's gray, then it can't be white," which ostensibly express *worldly* and specifically *metaphysical* knowledge, as really functioning to express the scorekeeping principles that determine the discursive significance of sentences of the form "x is gray" and "x is white." This precisely spells out a version of what has been called "modal normativism," a position originally charted out by Sellars (1953) and developed and defended most recently by Amie Thomasson (2020). This modal normativist account of conditionals of the above sort will enable us to precisely reconstruct the "worldly" entities that figure in intra-worldly and extra-worldly semantics—things like properties and possible worlds—as reifications of linguistic rules.

5.2 Modal Normativism and Logical Expressivism

In this dissertation, my main focus has been on what we might call "metaphysical structure" of the world. This structure includes, for instance, the fact that the various properties that things in the world might instantiate stand in the modal relations to one another that they do, entailing or being incompatible with one another. For instance, the property of being gray is incompatible with the property of being white in the sense that it's not possible for something to instantiate the property of being gray and also instantiate the property of being white, or, to put it differently, if something instantiates the property of being gray, it's not possible for it to instantiate the property of being white. To state another modal relation between properties, if something is black and something else is white, then necessarily, the first thing is darker than the second. These modalized conditionals articulate the metaphysical structure that I've claimed, in Chapter Three, is actually *constitutive* of these properties. In this chapter, I spell out a "modal normativist" view, according to which these conditionals are understood as expressing the norms governing the use of the predicates "gray," "white," "darker than," and so on. More precisely, on this framework, these conditionals are understood as expressing the scorekeeping principles that determine the semantic significance of these predicates.

The version of modal normativism defended here is owed most directly to Sellars (1958) and developments of Sellars by Brandom (2008, 2015). Recently, however, it has been notably defended by Amie Thomasson (2020) who argues particularly that the modal claims made in metaphysics are best understood on the normative expressivist model. On Thompson's account, a large class of disputes in metaphysics where the crucial claims being made are modal ones, are to be understood as really a kind of semantic dispute, where what is at issue is precisely the semantic norms governing the use of linguistic expressions. What Thomasson either doesn't realize or simply doesn't bring out is the radical consequences that modal normativism has for semantic theorizing. As I argued in the first three chapters, contemporary semantic theories generally take it, either explicitly or implicitly, that we can appeal to speakers' knowledge of these metaphysical modal relations in accounting for their knowledge of meaning. If these metaphysical modal

relations are really a "hypostatization," as Thomasson puts it, of the norms governing the use of linguistic expressions, and knowledge of these relations is really just a reflection of semantic knowledge, then any account that attempts to explain speakers' knowledge of meaning as depending on knowledge of these relations has things backwards. This, I've argued, is a fatal problem for the vast majority of contemporary semantic theories, insofar as they aim to explain speakers' knowledge of meaning. The current task is to explicate how the alternate semantic theory that I've laid out—discursive role semantics—is able to underwrite a thoroughgoing modal normativism.

As Thomasson proposes to spell out modal normativism, a (metaphysically) modalized sentence of the form "Necessarily φ" is true just in case φ is an object-language expression of an actual semantic rule or follows from such rules (8). Now, semantic rules are paradigmatically of conditional form, for instance: If you say, "*a* is black," then you can't say "*a* is white." Accordingly, the specific type of modalized expressions I'll principally concern myself with here are modalized *conditionals*, where the relevant conditionals that express semantic rules are, even if lacking an explicit modal operator, still understood as *implicitly* modal. Of course, the idea that conditionals even lacking explicit modal operators are still often implicitly modal in an important sense is a familiar one, spelled out perhaps most influentially in the work of Kratzer (1978, 1979, 1981). The approach to conditionals taken here, however, is quite different, aligning more directly with a *logical expressivist* account of conditionals, developed most influentially by Brandom (1994, 2008, 2018). According to Brandom, conditionals play the fundamental expressive role of enabling us to make explicit relations of consequence that determine the semantic significance of ordinary, non-logical expressions. With this notion of "consequence" understood, pragmatically in terms of scorekeeping principles, the relevant notion of consequence principally expressed by the conditional is that of *committive* consequence.[1] Thus, a conditional of the

[1]Note, that, in this context, the notion of committive consequence explicated here corresponds to what Brandom sometimes treats as the principle notion of consequence definable from his framework: incompatibility entailment (1994, 160; 2008, 117-175). p incompatibility entails q just in case every set of sentences incompatible with q is incompatible with p. Generalizing this notion of incompatibility entailment, we might put it in the following terms:

Incompatibility Entailment: p incompatibility entails q if, for all Γ, if $\Gamma \vdash \ominus_\alpha\langle q\rangle$, then $\Gamma \vdash \ominus_\alpha\langle p\rangle$.

Going from committive consequence to incompatibility entailment, if q is a committive consequence of p, we have $\oplus_\alpha\langle p\rangle \vdash \oplus_\alpha\langle q\rangle$. By Reversal we have $\ominus_\alpha\langle q\rangle \vdash \ominus_\alpha\langle p\rangle$, and, by (Simple) Transitivity, we have, for any

form $\varphi \rightarrow \psi$ expresses that commitment to φ commits one to ψ. Given our definition of commitment to a negation in terms of preclusion of entitlement to the negated sentence, we will also want to say that a conditional with a negated consequent of the form $\varphi \rightarrow \neg\psi$, though it directly expresses a relation of committive consequence (that commitment to φ commits one to $\neg\psi$), indirectly expresses an underlying relation of preclusive consequence: that commitment to φ *precludes entitlement* to ψ.

Though Thomasson draws her inspiration from Sellars in developing modal normativism, the use to which Sellars actually puts modal normativism in his philosophical theorizing, and the use to which it will be put here, is much more radical than that to which Thomasson puts it. While Thomasson claims that distinctively *modal* properties, such as the property of being necessarily incompatible with the property of being white, possessed by the property of being black, are reifications of linguistic rules, she never makes the claim that even *non-modal* properties, such as the property of being black itself, are likewise reifications of linguistic rules. That is the claim we'll make here. On the account we'll develop, the property of being black, for instance, *just is* that bit of metaphysical structure articulated by the set of modalized conditionals that express the scorekeeping principles governing the use of the predicate "black." This is not to say that the property of being black is necessarily a *mere* reflection of linguistic rules—this bit of metaphysical structure may well be instantiated by extra-linguistic reality. We will consider this possibility in the next chapter. The aim of this chapter, however, is to articulate an account of properties as reifications of discursive roles that doesn't presuppose worldly knowledge, thus not falling prey to the form of the Myth that plagues worldly semantics. The logic of conditionals explicated in the section following next, owed to Mark Lance and Philip Kremer (1994), makes these ideas precise. Before turning to that logic, however, let me first briefly distance the approach to be taken here from the approach to logical expressivism that has been taken by Brandom and his collaborators in recent years.

set of normative positions Γ, such that $\Gamma \vdash \ominus_\alpha\langle q\rangle, \Gamma \vdash \ominus_\alpha\langle p\rangle$. Going in the other direction, if p incompatibility entails q, we have, by CO $\ominus\langle q\rangle \vdash \ominus\langle q\rangle$, by the incompatibility entailment, we have $\ominus\langle q\rangle \vdash \ominus\langle p\rangle$, and, by Reversal, $\oplus\langle p\rangle \vdash \oplus\langle q\rangle$. Note that we no longer have this convergence in notions if we go substructural. For instance, commitment to "Sadie's a platypus" commits one to "Sadie's a mammal," but it's not the case that everything incompatible with "Sadie's a mammal" is incompatible with 'Sadie's a platypus," as "Sadie lays eggs" is (defeasibly) incompatible with the former, but not the latter. I take it that this is one of the main reasons why Brandom has stopped using the notion of incompatibility entailment in recent work.

5.3 The ROLE Approach to Conditionals

Logical expressivism, in its formal details, has been developed most substantially by members of the Research on Logical Expressivism (ROLE) working group, led by Robert Brandom and Ulf Hlobil, and whose principle members also include Daniel Kaplan, Shuhei Shimamura, Rea Golan, and myself.[2] In a series of papers (Hlobil 2016, Shimamura 2017, Hlobil 2017, Kaplan 2018, Hlobil 2018, Brandom 2018, Shimamura 2019) and unpublished work, members of this group have formally developed a conception of expressivism originally put forward by Brandom (2008), putting forward a general program for logical expressivism and various specific implementations of it: various specific "expressivist logics" designed to function in this formal expressivist framework. Though the broader formal setting adopted here is quite different, as will be made clear shortly, the account of specifically logical vocabulary provided in the previous chapter can be seen as of belonging to this general program, and the particular bilateral sequent calculus, can be seen as a particular implementation of it. Indeed, the sequent system provided there is equivalent to the main sequent system proposed by the group, the multiple conclusion sequent calculus NM-MS (Non-Monotonic Mulit-Succident), originally proposed by Kaplan (2017). As explained in the previous chapter, the bilateral sequent calculus has one crucial advantage over its multiple-conclusion twin: because it is a single conclusion sequent calculus, we can understand the sequents that figure it in terms of their function to update scorecards. All of this is essentially in line with the ROLE approach, and, indeed, owes itself to it. The difference in the formal framework developed here, however, becomes clear when it comes to conditionals (or, at least, those that are particularly pertinent to the project here), and this requires diverging from the ROLE approach.

The ROLE approach proceeds along the following lines. We start with atomic language $\mathcal{L}_0$, sentences of which are related by a material *base consequence* relation $\vdash_0$. We then construct a sequent calculus that *extends* this atomic language to a logically complex language $\mathcal{L}$ that includes sentences containing, for instance, the expressions "$\rightarrow$" and "$\wedge$," which related by an *extended consequence* relation $\vdash$. To see how such a calculus can be

[2] See https://logicalexpressivism.wixsite.com/role

thought of as an "expressivist logic," suppose our base consequence relation $\vdash_0$, relating sentences of $\mathcal{L}_0$, contains the following sequent:

a is gray, b is white $\vdash_0$ a is darker than b

In this context, such a sequent is understood not as expressing a scorekeeping principle, but, rather as expressing an implication relation that underwrites the inferences speakers of $\mathcal{L}_0$ make. So, speakers of $\mathcal{L}_0$ infer "a is darker than b" from both "a is gray" and "b is white." However, they don't have any way of making this inferential relation explicit in the form of a claim. Now consider the expressive capacity of speakers of $\mathcal{L}_1$, an extension of $\mathcal{L}_0$ that includes sentences containing "$\rightarrow$" and "$\wedge$." We might think of these speakers of $\mathcal{L}$ as upgraded speakers of $\mathcal{L}_0$, speakers of $\mathcal{L}_0$ who now comprehend the inferential significance of a sentence of $\mathcal{L}$ in virtue of having their inferential capacities algorithmically expanded by way of the following rules (Hlobil, 2016; Brandom 2018):

$$\frac{\Gamma \vdash \varphi \quad \Gamma \vdash \psi}{\Gamma \vdash \varphi \wedge \psi} \wedge_R \qquad \frac{\Gamma, \varphi, \psi \vdash \chi}{\Gamma, \varphi \wedge \psi \vdash \chi} \wedge_L \qquad \frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \rightarrow \psi} \rightarrow_R$$

Unlike speakers of $\mathcal{L}_0$, speakers of $\mathcal{L}$ *do* have a way of making the inferential relation that obtains between "a is gray" and "b is white" and "a is darker than b" explicit. For, $\mathcal{L}$ contains the sentence 'If a is gray and b is white, then a is darker than b," and this sentence, in the extended consequence relation, follows from the empty set in virtue of the following derivation:

$$\frac{\dfrac{\text{"}a\text{ is gray," "}b\text{ is white" } \vdash \text{"}a\text{ is darker than }b\text{"}}{\text{"}a\text{ is gray and }b\text{ is white" } \vdash \text{"}a\text{ is darker than }b\text{"}}\ L\wedge}{\vdash \text{"If }a\text{ is gray and }b\text{ is white, then }a\text{ is darker than }b\text{"}}$$

Since this sentence follows from the empty set of sentences, it might be thought of as a "material tautology:" a sentence that is assertable in virtue of the material consequence relation alone. It it thus a claim that can be made in $\mathcal{L}$—something speakers of $\mathcal{L}$ can *say*—that *makes explicit* an inferential norm (the goodness of inferring "a is darker than b" from "a is white" and "b is gray") that was only *implicit* in what speakers of $\mathcal{L}_0$ *did*.

To take the ROLE approach to conditionals here would be to introduce them in just the way that we have introduced rules for the other connectives. For instance, in our bilateral

set-up, the natural way to introduce a conditional is to define it by way of the following rules:

$$\frac{\Gamma, \oplus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle}{\Gamma \vdash \oplus_\alpha\langle\varphi \to \psi\rangle}\ \oplus_{\to} \qquad\qquad \frac{\Gamma \vdash \oplus_\alpha\langle\varphi\rangle \quad \Gamma \vdash \ominus_\alpha\langle\psi\rangle}{\Gamma \vdash \ominus_\alpha\langle\varphi \to \psi\rangle}\ \ominus_{\to}$$

These rules define the material conditional. That is, it comes out, according to them, that being committed to a claim of the form $\varphi \to \psi$ is the same as being committed to $\neg(\varphi \wedge \neg\psi)$ or, equivalently, $\neg\varphi \vee \psi$. While the language can indeed be extended to accommodate the material conditional in this way, and, indeed, it can be useful to do so, there is reason to want to model the rules governing the use of conditional expressions in a somewhat different way, at least insofar as we want to think about conditionals as expressing scorekeeping principles. It's not hard to see that there is something problematic about these rules, given the interpretation of the signs and the turnstile that has been developed here. For instance, the preclusive conditional rule, which wears the materiality of the conditional it defines on its sleeve, is obviously problematic, from an expressivist perspective as it seems to require far to much in order to be precluded from being entitled to a conditional. To show how deep the issue here is, however, let us focus on the positive conditional rule, which Brandom (2018), though explicitly a pluralist about conditionals, claims is the minimal requirement for something's counting as a conditional at all.

Let us first note that the structural principle of Containment (CO) gives us the sequent $\oplus\langle q\rangle, \oplus\langle p\rangle \vdash \oplus\langle q\rangle$, for any sentences p and q, and so an application positive conditional rule gives us the result that, for any sentences p and q, $\oplus\langle q\rangle \vdash \oplus\langle p \to q\rangle$. On this framework, having this scorekeeping principle would amount to scoring anyone who we score as committed to q as committed to the conditional $q \to p$, and thus, on the interpretation of conditionals as expressing scorekeeping principles, as committed to the scorekeeping policy of scoring anyone who's committed to p to be committed to q. That, of course, seems like a very bad result. After all, why should anyone who's committed to q take anyone who's committed to some irrelevant p to be committed to q? Such a policy would actually preclude someone from taking anyone to be really disagreeing with them, taking anyone, regardless of their commitments, to be committed to just what one is committed

to oneself! Moreover, though one more application of the positive conditional rule, we get that anyone, regardless of their commitments, is committed to $p \to (q \to p)$. Thus, for instance, anyone would be taken to be committed to "If *a* is black, then if *a*'s white, then *a*'s black." Not only does this sound terrible, but thinking of conditionals as expressing scorekeeping principles enables us to makes sense of why it does: commitment to "*a* is black" does not bring with it commitment to scoring anyone who one scores as "*a* is white" as committed to "*a* is black." Of course, these are just the paradoxes of material implication, understood in this scorekeeping setting. I'm just bringing it out to note how bad they seem in this context.

Now, it is possible to try to resolve this issue by going relevant in some way or another. For instance, following Shimamura's (2017) proposal, we might consider just conditionals in relevant regions of the consequence relation that satisfy only *Reflexivity* and not CO, thus ruling out the sequent $\oplus_\alpha\langle p\rangle, \oplus_\alpha\langle q\rangle \vdash \oplus_\alpha\langle p\rangle$. The real problem here, however, is clearly not CO, which is trivially correct on the interpretation of the turnstile that has been laid out here. Rather, the problem here is the positive conditional rule, which is a form of the *Deduction Theorem*. Considering the standard formulation of it in an unsigned system, the Deduction Theorem is the following principle:

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi}$$

The idea is that if, relative to a background set Γ, one can derive ψ from φ, then, relative to Γ, one can derive $\varphi \to \psi$. Brandom (2019) and Hlobil (2017) have argued that, insofar as "$\vdash$" signifies a relation of implication, then any conditional that can rightly be said to *express* that implication relation must support the Deduction Theorem. And this is presumably correct in the context of the ROLE approach explicated above. There is, however, a basic problem with the Deduction Theorem, insofar as it's interpreted in the framework proposed here.

On the framework proposed here, a sequent of the form $\oplus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle$ is understood as expressing, in the metalanguage, the principle of scoring anyone who's committed to φ to be committed to ψ. Insofar as conditionals are understood as expressing, in the object language, such scorekeeping principles, then, presumably, commitment to a conditional

should be treated as commitment to a scorekeeping principle. Thus, a sequent of the form $\vdash \oplus_\alpha\langle\varphi \to \psi\rangle$ would express the principle of scoring anyone, regardless of what they're committed to, to be committed scoring anyone who's committed to φ to be committed to ψ. Moving from $\oplus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle$ to $\vdash \oplus_\alpha\langle\varphi \to \psi\rangle$, as the deduction theorm lets us, amounts to a scorekeeper projecting their scorekeeping principles upon everyone else, taking it that if *they* keep score in a certain way *themself*, then so must *everyone else*. Thus, the failure of the deduction theorem, on this framework, is the result of the cross-perspectival interplay of "$\vdash$" and "$\to$". In the context of a scorekeeping principle that we hold, "$\vdash$" is the location we use to think of *our own* scorekeeping principles, whereas "$\to$" is the locution we use to think about the scorekeeping principles of *other scorekeepers*. In an expression of the form "$\oplus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi \to \chi\rangle$," the "$\vdash$" is expressing *our* scorekeeping principle, whereas the "$\to$" is expressing the scorekeeping principle of an arbitrary *other scorekeeper*. The reason why the move from $\Gamma, \oplus_\alpha\langle\varphi\rangle \vdash \oplus_\alpha\langle\psi\rangle$ to $\Gamma \vdash \oplus_\alpha\langle\varphi \to \psi\rangle$ is not a good one, on this way of thinking, is because it does not at all follow from *my* scoring anyone who is committed to Γ along with φ to be committed to ψ to my scoring anyone who is committed to Γ to be *themselves* committed to scoring *someone else* who is committed to φ to be committed to ψ.

Now, I should be clear, I am not arguing that there is anything wrong with the ROLE approach per se. The ROLE simply involves an abstraction from the perspectival landscape in which contents are conferred, concerning itself just with the non-perspectival structure of the contents conferred by a discursive practice (or perhaps with the mono-perspectival structure of inferring in the way those contents compel one to infer) rather than the multi-perspectival structure of the discursive practice that actually confers those contents. While this is surely a worthwhile elucidatory project of inferentially explicating semantic contents (and, moreover, explicating how those contents can be inferentially explicated), my aim here, as I take the aim of Brandom's *Making It Explicit* to have been, is the bolder project of actually *accounting* for these contents in terms of the structure of the discursive practice that confers them. As Brandom makes clear, such a practice essentially involves the existence of multiple perspectives from which the conceptual contents conferred by that practice can be articulated. Indeed, only by appreciating the way in which, as Brandom (1994) puts it, "conceptual contents are *essentially expressively perspectival*"

(590), can we make sense of those contents as being *objective*, as concerning things that are what and how they are, independently of what or how we take them to be. This notion of objectivity comes into view, in the first instance, by thinking of the commitments that one undertakes as such as to be attributed to *oneself* from the perspective of *someone else* who has a different set of scorekeeping principles than oneself. Only in virtue of this potential perspectival distinction can one think of the commitments that one *really* undertakes, in making some claim, as potentially distinct from the set of commitments one *takes oneself* to undertake. A formal development of the perspectival account of objectivity offered in Chapter Eight of *Making It Explicit* is beyond the scope of the current project.[3] However, the multi-perspectivality on which that account is based is an absolutely essential feature of the framework proposed here. Scorekeeping principles must be such as to potentially vary from perspective to perspective, and so the conditionals that function to express scorekeeping principles must be sensitive to this potential variation. So, the ROLE approach, insofar as it is essentially non-perspectival or mono-perspectival, is simply incapable of making formal sense of conditionals that express scorekeeping principles.

At this point, one might be tempted to ask, what sequent rules *should* we give the conditional which *don't* support the deduction theorem? But I don't think that's the right question. Instead, I think we should ask, what other system should we use? Let me explain. We've modeled logical vocabulary thus far in terms of what commitments one undertakes and what entitlements one precludes oneself from in using that vocabulary. For instance, in using a sentence of the form $\varphi \wedge \psi$ one commits oneself to φ and one also commits oneself to ψ, and, in using a sentence of the form $\neg\varphi$, one precludes oneself from being entitled to φ. The function of conditional vocabulary, however, and, in particular, the sort of modalized conditionals that will concern us here, is not simply to *undertake* commitments but to express the scorekeeping principles in accordance with which one *attributes* commitments. Of course, expressing a scorekeeping principle in the form of a conditional claim *is* to undertake a commitment, but note that this involves a conceptual

[3]The account of objectivity developed in the next chapter, which draws on Brandom's (2008, 2019) later work along with the work of John Haugeland (1998), will ultimately have to be merged with the perspectival account of objectivity for a full account.

shift, from thinking about the undertaker of commitments *third-personally*, as being the one *on whom score is kept*, to thinking about the undertaker of commitments *first-personally*, the one *who is keeping score*, expressing the principles by which one does it. If one thinks of these conditionals as part of the "linguistic organ of semantic self-consciousness" (384), as Brandom wonderfully puts it, one will want an account of logical vocabulary that is articulated, at least in the first instance, *from the perspective of one who is semantically self-conscious*, who is capable of using conditional vocabulary to articulate the principles according to which they score other players as committed, entitled, and precluded from being entitled to various claims.

Of course, a complete account of conditionals will have to integrate the various perspectives that can be had with respect to the making of conditional claims. That is, it will not only have to specify the conditions under which a scorekeeper should commit themself to a conditional, but specify what happens when such a scorekeeper expresses that conditional commitment in a discursive practice, how other players update their scorecards in response. A proper account of these updates will require expanding the notion of a scorecard such that it not only keeps track of the normative positions that the various players occupy but also the different sets of scorekeeping principles that the various speakers have, which they express with the use of conditionals. This can and should be done, but I will not do it here. It is sufficient, for our purposes, to articulate an account of conditional claims from the perspective of a potential maker of those claims, a scorekeeper who is reflectively aware of their scorekeeping practices and who makes conditional claims accordingly. Such an account is what we need in order to make sense of the "worldly" knowledge with which we've been concerned here, articulating it as really a kind of semantic knowledge. Let us now turn to a formal system that will enable us to do just this.

5.4 Lance and Kremer's Commitment Logic

The way of introducing conditionals I will propose is owed to Mark Lance and Philip Kremer (1994), and it connects naturally to Lance's (1996) proposal for making sense of

quantifiers in inferentialist terms. On this approach, conditionals understood in terms of their function to express principles of committive consequence, as explicated through a Fitch-style natural deduction system through which we hypothetically attribute commitments to arbitrary players, and attribute the commitments that follow. In the basic case, we are to assert the conditional $\varphi \to \psi$ is, if, on the supposition that an arbitrary player α_1 is committed to φ, we score α_1 as committed to ψ. Asserting an embedded conditional of the form $\varphi \to (\psi \to \chi)$ amounts to the expressing the principle of scoring anyone who is committed to φ as committed to scoring anyone who is committed to ψ as committed to χ. To entitle ourselves to assert such a conditional, we assume an arbitrary player α_1 is committed to φ, and see if, given this supposition, they score another arbitrary player α_2 who they score as committed to ψ to be committed to χ. And so on for an arbitrary number of nestings. Natural rules avoid the paradoxes of material implication, making explicit the reasoning through which we've rejected the deduction theorem, and, even though the rules themselves don't involve any explicit talk of modality, the conditionals they define, with various tweaks, turn out to be the strict conditionals of familiar modal logics. I will principally consider just one of four possible systems that Lance and Kremer propose, the weakest one, which turns out to define the strict conditional of K, though I will briefly consider modifications resulting in stronger systems at the end of this section.

Changing the notation from Lance and Kremer slightly to align it with the notation employed here, and explicitly connecting the system to the general framework here, in the basic case, where we have a conditional that expresses some material scorekeeping principle we have, we'll have a proof of the following form:

1	$\oplus_{\alpha_1}\langle p\rangle$	asm.
2	$\oplus_{\alpha_1}\langle q\rangle$	PA (1, $\oplus_{\alpha}\langle p\rangle \vdash \oplus_{\alpha}\langle q\rangle$)
3	$p \to q$	$\to_I$ (1, 2)

So, for instance, if p is "a is crimson" and q is "a is red," then I can assert "If a is crimson, then a is red" just in case I score anyone who is committed to "a is crimson" to be committed to "a is red." To assert such a thing is to express a scorekeeping principle that I have, the one in virtue of which I score anyone who is committed to "a is crimson" to be committed to "a

is red." To make this explicit and connect the formal framework proposed in the previous chapter, in which each player has a set of scorekeeping principles π, I have added the following *Principle Application* (PA) rule illustrated in line (2) of the above proof:[4]

> Principle Application (PA): Given $\oplus_{\alpha_1}\langle\varphi_1\rangle, \oplus_{\alpha_1}\langle\varphi_2\rangle \ldots \oplus_{\alpha_1}\langle\varphi_n\rangle$ in the same subproof, if $\oplus_{\alpha}\langle\varphi_1\rangle \ldots \oplus_{\alpha}\langle\varphi_n\rangle \vdash \oplus_{\alpha}\langle\psi\rangle \in \pi$, infer $\oplus_{\alpha_1}\langle\psi\rangle$

The use of PA at line two in the proof above is an instance of this rule just in case we have the scorekeeping principle $\oplus_{\alpha}\langle p\rangle \vdash \oplus_{\alpha}\langle q\rangle$.

Let us now consider the conditional introduction and elimination rules proposed by Lance and Kremer. Lance and Kremer's proof theory begins with the thought that there can be nested attributions of commitments. So, not only can we think of an arbitrary player α_1 as being committed to φ, but we can also think of α_1 as being committed to the claim that some other arbitrary player α_2 is committed to φ. If α_1 is committed to the claim that α_2 is committed to φ, we can write this as "$\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$."[5] An embedded conditional of the form $(\varphi \to (\psi \to \chi))$ can then be understood as saying that anyone who is committed to φ is committed to the claim that anyone who is committed to ψ is committed to χ. So, to give a proof of this conditional, we'd start with the hypothesis $\oplus_{\alpha_1}\langle\varphi\rangle$ and set out prove $\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\chi\rangle\rangle$, given the further hypothesis $\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\psi\rangle\rangle$. The conditional introduction rule, generalizing this notion of a hypothetical proof to an arbitrary number of nestings, is given as follows:

> $\to_I$**:** Given a proof of $\oplus_{\alpha_1} \ldots \oplus_{\alpha_{n+1}}\langle\psi\rangle$ on hypothesis $\oplus_{\alpha_1} \ldots_{\alpha_{n+1}}\langle\varphi\rangle$, infer $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n}$ $\langle\varphi \to \psi\rangle$, where $n \geq 0$.

Given this introduction rule, there is a natural corresponding elimination rule, a form of modus ponens. The thought is that if α_1 is committed to $(\varphi \to \psi)$, then α_1 will score anyone they score as committed to φ to be committed to ψ. So, from $\oplus_{\alpha_1}\langle\varphi \to \psi\rangle$ and $\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$, we can assert $\oplus_{\alpha}\langle\oplus_{\alpha_2}\langle\psi\rangle\rangle$. The conditional elimination rule, generalizing this notion of a modus ponens to an arbitrary number of nestings, is given as follows:

[4]Though Lance and Kremer use examples such as the fact that commitment to "Fido is a Dog" commits one "Fido is a mammal" to motivate the proof theory for the conditional that they develop, they never actually provide a way of integrating such material relations of committive consequence. That is what this rule does.

[5]Note that the commitment sign, which previously only appeared in the *meta-language*, is not appearing in the *object language*, embedded in claims to which arbitrary scorekeepers are scored as committed.

$\rightarrow_E$: From $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \rightarrow \psi \rangle$ and $\oplus_{\alpha_1} \ldots \oplus_{\alpha_{n+1}} \langle \varphi \rangle$, infer $\oplus_{\alpha_1} \ldots \oplus_{\alpha_{n+1}} \psi$, where $n > 0$

Thus, if some scorekeeper is committed to a conditional of the form $\varphi \rightarrow \psi$, and they score someone as committed to ψ, then they'll score them as committed to φ.

Because of the logical rules provided in the previous chapter, the Principle Application rule, which is not restricted to material scorekeeping principles, will let us express, in the form of conditionals, both *material* consequences and *logical* consequences. So, any theorem of classical logic will be able to be expressed in the form of a conditional. However, because we may want to logically combine conditionals, so that we can express, for instance, the transitivity of consequence with a conditional of the form $((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \chi)) \rightarrow (\varphi \rightarrow \chi)$, it will be helpful to add in to this system rules for conjunction as well. Lance and Kremer propose the following natural rules for conjunction:

$\wedge_I$: From $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \rangle$ and $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \psi \rangle$, infer $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \wedge \psi \rangle$
$\wedge_{E_L}$: From $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \wedge \psi \rangle$, infer $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \rangle$
$\wedge_{E_R}$: From $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \wedge \psi \rangle$, infer $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \psi \rangle$

The conditional and the conjunction can be seen as playing a special expressive role since, jointly, they enable us to express the various *structural* scorekeeping principles (or, more precisely, any instance of a structural scorekeeping principle). For instance, for Transitivity, we have a proof such as the following:

1	$\oplus_{\alpha_1}\langle(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \chi)\rangle$	asm.
2	$\oplus_{\alpha_1}\langle(\varphi \rightarrow \psi)\rangle$	$\wedge_{E_L}$ (1)
3	$\oplus_{\alpha_1}\langle(\psi \rightarrow \chi)\rangle$	$\wedge_{E_R}$ (1)
4	$\quad\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$	asm.
5	$\quad\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\psi\rangle\rangle$	$\rightarrow_E$ (2, 4)
6	$\quad\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\chi\rangle\rangle$	$\rightarrow_E$ (3, 5)
7	$\oplus_{\alpha_1}\langle(\varphi \rightarrow \chi)\rangle$	$\rightarrow_I$ (4-6)
8	$((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \chi)) \rightarrow (\varphi \rightarrow \chi)$	$\rightarrow_I$ (1-7)

It's easy to construct similar proofs of all of the structural principles validated by the update semantics, as shown in the previous chapter (Sections 4.5 and 4.7), such as Monotonicity,

Permutation, Contraction, and so on. This logic is thus expressive in two ways. When the PA rule is used, conditionals express material or logical scorekeeping principles. When PA is not used, conditionals express (instances of) *structural* scorekeeping principles.[6]

Now that we have seen some of this system's expressive power, just with the rules for these two connectives, let us look at some of its interesting features. First, we should note that certain crucial paradoxes of material implication are avoided. For instance, $p \rightarrow (q \rightarrow p)$ is not provable in this system. The following "proof" fails to accord with the conditional introduction rule:

1	$\oplus_{\alpha_1}\langle p\rangle$	asm.
2	$\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle q\rangle\rangle$	asm.
3	$\oplus_{\alpha_1}\langle p\rangle$	reit. (1)
4	$\oplus_{\alpha_1}\langle q \rightarrow p\rangle$	$\rightarrow_I$ (2 – 3)? (fallacious step)
5	$p \rightarrow (q \rightarrow p)$	$\rightarrow_I$ (1-4)

In order for (4) to follow from (2) and (3), (3) would need to be $\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle p\rangle\rangle$. Only if we can show that α_1 scores anyone who they score as committed to q to be committed to p could we assert $\oplus_{\alpha_1}\langle q \rightarrow p\rangle$. Intuitively, even if α_1 is committed to p, it does not follow that α_1 scores anyone who is committed to q to likewise be committed to p. So, on this way of thinking about what is expressed by a conditional locution, $p \rightarrow (q \rightarrow p)$ should not be a logical truth. This is just the reasoning articulated above in connection with the ROLE approach to conditionals, made formally explicit, and this is the key contrast between this sort of system and the sort of systems considered in the context of the ROLE approach, in which $p \rightarrow (q \rightarrow p)$ straightforwardly follows from CO and the Deduction Theorem.

One fact about this system worth noting is that import/export does not hold. That is, $(\varphi \wedge \psi) \rightarrow \chi$ is not equivalent to $(\varphi \rightarrow \psi) \rightarrow \chi$. One can see this simply by noting that, while $p \rightarrow (q \rightarrow p)$ is not a theorem, $(p \wedge q) \rightarrow p$ is, as it can be proven quite simply as follows:

[6] In order to be able to express (instances of) *bilateral* scorekeeping principles, such as Reversal or Bilateral Reductio, we must introduce negation into this system as well. This can be done in a straightforward way, following the approach to bilateralism taken in the previous chapter. It's also worth noting that if we go substructural in our discursive role semantics, as I propose elsewhere, the rules for this sort of system will need to be modified in order to be properly expressive of that version of discursive role semantics.

1	$\oplus_{\alpha_1}\langle p \wedge q\rangle$	asm.
2	$\oplus_{\alpha_1}\langle p\rangle$	$\wedge_{E_L}$ (1)
3	$(p \wedge q) \rightarrow p$	$\rightarrow_I$ (1-2)

Intuitively, this distinction comes down to the fact that, whereas, clearly, everyone who is committed to $p \wedge q$ is committed to p, it's not the case that everyone who is committed to p is committed to scoring anyone who's committed to q to be committed to p. Though this makes a lot of intuitive sense in the current setting, this may appear to be a negative thing when we consider some concrete examples. Consider, for instance, that, though "If a is black and b is white, then a is darker than b" will be a theorem, "If a is black, then if b is white, then a is darker than b" will not be. The latter conditional comes out as saying that anyone committed to "a is black" is committed to scoring anyone who they score as committed to "b is white" to be committed to "a is darker than b," and that is not the case; one will only score such a person as committed to "a is darker than b" if one scores them as committed to "a is black" as well (or some other claim that, given "b is white," entails that "a is darker than b"). If our aim was to adequately represent the logic of natural language bare indicatives, this would of course be a very bad result. However, these conditionals are not really bare indicatives at all, but *strict* conditionals, more aptly expressed in natural language by modalized indicatives. Thus, a more apt natural language translation of $Ba \rightarrow (Wb \rightarrow Dab)$ of "Necessarily, if a is black, then, necessarily, if b is white, then a is darker than b," and this need not be interpreted as true, since it's not the case that a is necessarily black.[7]

Now, just as I'm not claiming that this account is an empirically adequate account of natural language bare indicatives, I'm also not claiming that it's an empirically adequate account of modalized indicatives of the form "Necessarily, if φ, then ψ." The aim here has been to spell out an account of the specific type of conditional that functions to express scorekeeping principles, and there is no guarantee that such a conditional precisely

[7]It's worth noting that if one wants a formal correlate of "If a is black, then if b is white, then a is darker than b," where this functions to express a scorekeeping principle we actually have, one can take the "If... then"s to be ambiguous between the modalized conditional of committive consequence and the material conditional, thus taking the sentence to be of the form $Ba \rightarrow (Wb \supset Dab)$. This is indeed derivable, given the rules provided here and the material conditional rules provided in the previous section.

corresponds to a natural language expression. The conditional defined does however, correspond to a formal language expression: the strict conditional of the modal logic K. As Lance and Kremer point out, the purely logical fragment of this system—which we get here by removing the use of the Principle Application rule—is identical to the conjunction and strict conditional fragment of K (1994, 383). So, this set of rules defines a perfectly tractable logic for the conditional, and, though the system through which the conditional was defined did not involve any explicit talk of modality, the resulting framework is that of a modalized conditional. Moreover, as Lance and Kremer show, the strict conditional of stronger modal logics result from modifying the conditional elimination rule. For instance, if we modify it so as to allow reasoning from $\oplus_{\alpha_1}\langle\varphi \rightarrow \psi\rangle$ and $\oplus_{\alpha_1}\langle\varphi\rangle$ to $\oplus_{\alpha_1}\langle\psi\rangle$, we get the strict conditional of T. Lance and Kremer propose four distinct systems, resulting from four distinct formulations of the conditional elimination rule (1994, 383). Though Lance and Kremer take no stand on which of the logics they propose is the correct logic of committive consequence, arguing for or against various principles concerning committive consequence, and thereby determining which modal logic has a privileged expressive role, is a formally tractable project in the foundations of metaphysical modality for the modal normativist to undertake. Actually undertaking it is well beyond the scope of the dissertation, however, and I will settle, for my purposes here, on the intuitiveness of the rules that define this system, which Lance and Kremer call C1.[8]

5.5 Quantifiers

One of the features of this proof system is that we can directly import standard natural deduction rules for the quantifiers, as explicated in inferentialist terms by Lance (1996). The form of the universal quantifier introduction rule that I will use is the following:[9]

[8] It is also worth noting that, in a subsequent paper, Lance and Kremer (1996) propose four systems of *relevant* committive consequence, where a conditional of the form $\varphi \rightarrow \psi$ is assertable only if commitment to φ is relevant to commitment to ψ. Such systems are also certainly worth exploring.

[9] A rule of this sort is proposed by MacFarlane (2021). The specific formulation of this rule is drawn from Garson's (2014, 17-20) rule for the box of modal logic. Any standard rule for the quantifier will do, but I use this version, which requires an explicit subproof, for extra conceptual clarity.

$$\begin{array}{|l} \boxed{a} \\ \hline \vdots \\ \Phi(a) \\ \forall x(\Phi(x)) \qquad \forall_I \end{array}$$

Where a doesn't occur in Φ

Where a line with a boxed name occurs, it introduces a subproof with a restriction on the reiteration rule: no previous line of the proof in which that name occurs can be reiterated into that subproof. This guarantees that the name, as it occurs in this subproof, functions *arbitrarily*. The intuitive idea behind this rule, is that, if we can show that some predicate Φ holds of *something* a, where we know *nothing* about a, then we will have thereby shown that Φ holds of *everything*. The ability to speak of "everything" here enables us to introduce a new kind of expression: the variable. The x, as it occurs in a sentence of the form "For all x, $\Phi(x)$" is not a *name* like a, but, rather, something that functions quite differently. Rather than functioning to *pick out* some particular thing, it *ranges* over everything.

Before turning to particular proofs in this system that use this rule, let us look at the general form of a proof that uses this quantifier rule:

$$\begin{array}{|l} \boxed{a} \\ \hline \quad \begin{array}{|ll} \oplus_{\alpha_1}\langle Fa\rangle & \text{asm.} \\ \hline \quad \vdots & \\ \quad \oplus_{\alpha_1}\langle Ga\rangle & \end{array} \\ Fa \rightarrow Ga \qquad \rightarrow_I \\ (\forall x)(Fx \rightarrow Gx) \qquad \forall_I \end{array}$$

Here, we begin our proof with a subproof that ensures that a functions as an arbitrary name. We then suppose that α_1 is committed to Fa. Now, there the proof goes on for some length and we are able to conclude that α_1 is committed to Ga. So, by our conditional introduction rule, we are able to conclude $Fa \rightarrow Ga$. Since this occurs within a subproof in which a functions as arbitrary, we can use the universal quantifier introduction rule and assert $\forall x(Fx \rightarrow Gx)$. On this construal, saying "Everything that's an F is a G" is a way of

expressing a principle of scoring anyone who's committed to a sentence of the form Fx to be committed to a sentence of the form Gx, where this is understood in terms of the fact, that when someone is scored as committed to the sentence Fa, where a is an arbitrary name, they're scored committed to Ga.

Now let's turn to a proof of a universally quantified conditional that expresses a material scorekeeping principle that we can actually construct in this framework. Let us recall, first, the way that we have conceived of scorekeeping principles relating *logically complex* sentences as generated from scorekeeping principles relating *atomic* sentences, which are, in turn, generated by scorekeeping principles relating sentence *frames*. Thus, the scorekeeping principle

$$\oplus_\alpha\langle a \text{ is gray}\rangle \vdash \oplus_\alpha\langle \neg a \text{ is white}\rangle$$

Is conceived of as generated from the following scorekeeping principle:

$$\oplus_\alpha\langle a \text{ is gray}\rangle \vdash \ominus_\alpha\langle a \text{ is white}\rangle$$

which is, in turn, is conceived as generated from the following scorekeeping principle:

$$\oplus_\alpha\langle x \text{ is gray}\rangle \vdash \ominus_\alpha\langle x \text{ is white}\rangle$$

by way of the following rule:

$$\frac{\Gamma, {}^{\oplus}\!/_{\ominus_\alpha}\langle\Phi_1(x_i)\rangle \ldots {}^{\oplus}\!/_{\ominus_\alpha}\langle\Phi_n(x_i)\rangle \vdash {}^{\oplus}\!/_{\ominus_\alpha}\langle\Psi(x_i)\rangle}{\Gamma, {}^{\oplus}\!/_{\ominus_\alpha}\langle\Phi_1(\tau)\rangle \ldots {}^{\oplus}\!/_{\ominus_\alpha}\langle\Phi_n(\tau)\rangle \vdash {}^{\oplus}\!/_{\ominus_\alpha}\langle\Psi(\tau)\rangle}$$

The significance of this mechanism of generating scorekeeping principles on sentences from scorekeeping principles on frames is that, if we have a scorekeeping principle on frames, we can generate a scorekeeping principle on sentences in which *any* singular term is substituted in for the variable. Thus, we can reason as follows:

1	$\boxed{a}$	
2	$\oplus_{\alpha_1}\langle a \text{ is gray}\rangle$	asm.
3	$\oplus_{\alpha_1}\langle \neg(a \text{ is white})\rangle$	PA (1, from $\oplus_\alpha\langle Gx\rangle \vdash \ominus_\alpha\langle Wx\rangle$)
4	a is gray $\rightarrow \neg(a$ is white$)$	$\rightarrow_I$ (1-4).
5	$(\forall x)(x$ is gray $\rightarrow \neg(x$ is white$))$	$\forall_I$ (1-5)

Thus, because we score an arbitrary agent who we take to be committed to "a is gray," for an arbitrary name a, to be precluded from being entitled to "a is white" and so committed to "It's not the case that a is white," we can assert, for an arbitrary a, "If a is gray, then it's not the case that a is white," and that enables us to assert "For all x, if x is gray, then it's not the case that x is white." Essentially, what this set of rules does, at least as it pertains to scorekeeping principles generated from principles on frames, is enable one to recover the quantificational vocabulary that was originally used to specify the method for generating scorekeeping principles through this substitution rule. The basic idea is that speakers' scorekeeping practices are *implicitly* universally quantificational, and what quantificational vocabulary in the object language does is enable them to make this feature of their scorekeeping practices *explicit*.

Let us look at a case in which we have embedded quantifiers. Consider, for instance, "Necessarily, if something's black, then there's nothing darker than it," or, in our quasi-formal language, $(\forall x)(x \text{ is black} \rightarrow (\forall y)(\neg y \text{ is darker than } x))$. We saw in the last chapter that we have the following material scorekeeping principle (on frames):

$$\oplus_\alpha\langle x \text{ is black}\rangle \vdash \ominus_\alpha\langle y \text{ is darker than } x\rangle$$

With this scorekeeping principle, we can reason as follows:

1	$\boxed{a}$	
2	$\quad \oplus_{\alpha_1}\langle a \text{ is black}\rangle$	asm.
3	$\quad \boxed{b}$	
4	$\quad \oplus_{\alpha_1}\langle a \text{ is black}\rangle$	reit.
5	$\quad \oplus_{\alpha_1}\langle \neg b \text{ is darker than } a\rangle$	PA (1, $\oplus\langle Bx\rangle \vdash \ominus_\alpha\langle Dyx\rangle$)
6	$\quad \oplus_{\alpha_1}\langle \forall y(\neg y \text{ is darker than } a)\rangle$	$\forall_I$ (3-5)
7	$a \text{ is black} \rightarrow \forall y(\neg y \text{ is darker than } a)$	$\rightarrow_I$ (2-6)
8	$\forall x(x \text{ is black} \rightarrow \forall y(\neg y \text{ is darker than } x))$	$\forall_I$ (1-7)

One more feature of the system is necessary in order for it to be properly expressive of the semantics put forward in the previous chapter. We should want to be able to say such things as "Necessarily, everything is the same shade as itself," and, thus far, we are

not able to, since we are only able to assert conditionals. To do this, we simply add the following rule.

> $\oplus_E$: From $\oplus_{\alpha_1}\langle\varphi\rangle$, occurring unembedded under any suppositions attributing commitments to α_1, infer φ

Thus, if we score an arbitrary agent as committed to some sentence φ, without supposing any other commitments on the part of that agent, we can simply assert that sentence. So, we can reason as follows:

1	$\boxed{a}$	
2	$\oplus_{\alpha_1}\langle a$ is the same shade as $a\rangle$	PA ($\vdash \oplus_\alpha\langle Sxx\rangle$).
3	a is the same shade as a	$\oplus_E$ (2)
4	$\forall x(x$ is the same shade as $x)$	$\forall_I$ (1-3)

Note that this system is one that functions to provide object-language expressions of semantic rules, or that which follows from them, and so, by Thomasson's modal normativist proposal, we can add "Necessarily" to a theorem we derive, even if that expression is not one of the modalized conditionals.

5.6 Reconstructing Intra-Worldly Semantics

In Chapter Three, we proposed the following definition of the property of being black, after considering several failed attempts at a definition:

> **⟦black⟧** = the property of being black =
>
> > The property such that, if something instantiates it, then, necessarily, it is darker than anything gray or white, nothing is darker than it, everything is either the same shade as it or lighter than it, and so on.

I claimed, in Chapter Three, that the "metaphysical structure" articulated by this definition was really nothing but a reification of the semantic norms governing the use of the predicate "black." I promised, when I proposed this definition in Chapter Three, that, by the end of Chapter Five, we would have the tools to think of this definition as, though ostensibly articulating a bit of metaphysical structure, as really functioning to express these norms. We can now do just that.

The first conditional, "If something's black, then, necessarily, it's darker than anything gray or white" can be put as follows:[10]

$\forall x \forall y((Bx \wedge (Gy \vee Wy)) \rightarrow Dxy)$

First, we derive the logically complex scorekeeping principle (on frames) $Bx \wedge (Gy \vee Wy) \vdash Dxy$ from the atomic scorekeeping principles $\oplus\langle Gy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle$ and $\oplus\langle Wy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle$ as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\oplus\langle Gy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle}{\ominus\langle Dxy\rangle, \oplus\langle Bx\rangle \vdash \ominus\langle Gy\rangle}\,\text{RV} \quad \dfrac{\oplus\langle Wy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle}{\ominus\langle Dxy\rangle, \oplus\langle Bx\rangle \vdash \ominus\langle Wy\rangle}\,\text{RV}}{\ominus\langle Dxy\rangle, \oplus\langle Bx\rangle \vdash \ominus\langle Gy \vee Wy\rangle}\,\ominus_{\vee}}{\ominus\langle Dxy\rangle \vdash \ominus\langle Bx \wedge (Gy \vee Wy)\rangle}\,\ominus_{\wedge}}{\oplus\langle Bx \wedge (Gy \vee Wy)\rangle \vdash \oplus\langle Dxy\rangle}\,\text{RV}$$

We can then use the substitution rule to get $\oplus\langle Ba \wedge (Gb \vee Wb)\rangle \vdash \oplus\langle Dab\rangle$. Now, given the PA rule, we can derive the universally quantified conditional used in this definition of the property of being black as follows:

1	$\boxed{a}$	
2	$\quad\boxed{b}$	
3	$\quad\quad\oplus_{\alpha_1}\langle Ba \wedge (Gb \vee Wb)\rangle$	asm.
4	$\quad\quad\oplus_{\alpha_1}\langle Dab\rangle$	PA (3)
5	$\quad(Ba \wedge (Gb \vee Wb)) \rightarrow Dab$	$\rightarrow_I$ (3-4)
6	$\forall y((Ba \wedge (Gy \vee Wy)) \rightarrow Day)$	$\forall_I$ (2-5)
7	$\forall x \forall y((Bx \wedge (Gy \vee Wy)) \rightarrow Dxy)$	$\forall_I$ (1-6)

This makes clear the way in which this universally quantified conditional really functions to express certain material scorekeeping principles, in particular, these ones:

$\oplus_\alpha\langle Wy\rangle, \oplus_\alpha\langle Bx\rangle \vdash \oplus_\alpha\langle Dxy\rangle$
$\oplus_\alpha\langle Gy\rangle, \oplus_\alpha\langle Bx\rangle \vdash \oplus_\alpha\langle Dxy\rangle$

We've already seen how the second universally quantified conditional in this definition, "If something's black, then, necessarily, there's nothing darker than it," can be understood as functioning to express the following material scorekeeping principle:

[10] As mentioned in the note above, if one doesn't like this formulation, and one prefers a formulation with an embedded quantifier, one can, with the material conditional, formulate this as follows: $\forall x((Bx \rightarrow \forall y((Gy \vee Wy) \supset Dxy)$

$$\oplus_\alpha\langle Bx\rangle \vdash \ominus_\alpha\langle Dyx\rangle$$

And if we cashed out that "and so on," we would eventually articulate all of the material scorekeeping in which the predicate "black" figures.

This account provides the formal cash for the Sellarsian claim, proposed in Chapter Three, that property of being black is a linguistic and conceptual *reification* of the norms governing the use of the predicate "black." We've, in effect, provided a system of transposing the scorekeeping principles of the normative framework provided in the previous chapter, which determine the normative significance of the use of the predicate "black," into the object language, as modalized conditionals which say, if something is black, what else must follow. The *structure* of these norms gets preserved through this transposition, but the modal *flavor* shifts. Whereas the modality that characterizes the scorekeeping principles has a *normative* flavor, the modality that characterizes the conditionals that express those scorekeeping principles has an *alethic* flavor. This just is the process of linguistic reification—the construction of a "thing" through the linguistic transposition from the normative to the alethic. The basic diagnosis of worldy semantics—its fundamental mistake—is its blindness to this reification process, taking the reifications of our linguistic norms to be self-standing worldly entities that can thereby function to explain those norms.

Let us briefly return to the arguments of Capplen and Lepore (2005), discussed in Chapter Three (Section 3.4). In defending the view that an account of what properties are is not a matter for semantics, Cappellen and Lepore crucially rely on the point that a claim that articulates what the property of being red is, for instance, "is not a claim about language; in particular, it is not a claim about the word 'red,"' (160). This is, indeed, a crucial point of the modal normativist position, made by Sellars, elaborated at length by Thomasson, and formally explicated by this technical account. The conditional that (in part) articulates what it is for something to be red, for instance "If something's red, then it must be colored," is not a a claim *about* the word "red," nor is it a claim about a scorekeeping principle normatively relating utterances of sentences containing the word "red." It *expresses* such a scorekeeping principle, but it's not *about* such a scorekeeping

principle. Rather, insofar as one thinks in the reified mode, thinking of conditionals like this as articulating the bits of metaphysical structure constitutive of properties, then this claim is a claim about the property of being red and the property of being colored, one that says of these properties that the former stands in an entailment relation to the latter. Cappelen and Lepore infer from the *correct* claim that metaphysical questions about the essences of properties "are not questions about language" to the *incorrect* claim that "they are nonlinguistic questions," concluding "Not only is there no reason to think these worries can be solved by doing semantics, there is no reason to think they have anything at all to do with semantics," (159). As we have shown, these questions, although not about language, have everything to do with language. Moreover, not only do they have everything to do with semantics, but they *can* actually be solved by doing semantics, for the answers to them just are the expressions of the scorekeeping principles that determine the semantic significance of the predicates whose worldly reifications the questions explicitly concern.

5.7 Reconstructing Extra-Worldly Semantics (and More)

Whereas properties are reifications of the norms governing the use of predicates, states of affairs are reifications of the norms governing the use of sentences. Just like the property of being black, the state of affairs consisting in a's being black, for instance, is to be understood in terms of the modal relations that this state of affairs bears to other states of affairs, for instance, excluding the state of affairs consisting in a's being white, and, if combined with the state of affairs consisting in b's being gray, including the state of affairs consisting in a's being darker than b, and so on. The modalized conditionals that articulate these modal relations between states of affairs express scorekeeping principles of committive and preclusive on sentences. In this context, consider Planatinga's (1976) conception of a possible world w as a maximal possible state of affairs. This is a state of affairs such that, for every state of affairs S, w either *includes* S or *excludes* S, and there's no state of affairs S such that w both includes and excludes S. It's easy to see that this is the worldly correspondent of a maximal, coherent, single-player scorecard: a scorecard σ, with scores kept for a single arbitrary player α_1, conforming to a set of material scorkeeping principles

π, such that, for every atomic sentence p, σ either contains $\oplus_{\alpha_1}\langle p\rangle$ or $\ominus_{\alpha_1}\langle p\rangle$, and there is no atomic sentence q such that σ contains both $\oplus_{\alpha_1}\langle q\rangle$ and $\ominus_{\alpha_1}\langle q\rangle$.

The new non-primitivist actualist conception of possible worlds (King 2007), where (non-actual) possible worlds are identified with maximal uninstantiated properties that the world could have instantiated, is simply a syntactically varied reification of this same notion of maximal, coherent, single-player scorecards. To say "The world could be such that *a* is gray, and it could be such that *a* is white, but it can't be such that *a* is both gray and white, since, if something's gray, it can't be white." is a way of expressing, in worldly vocabulary, that there is a coherent set of normative assignments that contains $\oplus_{\alpha_1}\langle a \text{ is gray}\rangle$ and a coherent set of assignments that contains $\oplus_{\alpha_1}\langle a \text{ is white}\rangle$, but no coherent set of assignments that contains both $\oplus_{\alpha_1}\langle a \text{ is gray}\rangle$ and $\oplus_{\alpha_1}\langle a \text{ is white}\rangle$, since commitment to a sentence of the form "*x* is gray" precludes entitlement to a sentence of the form "*x* is white." Note that the non-primitivst about worlds, unlike the primitivist, will be happy to say that the *reason* the world cannot be such that *a* is gray and *a* is white is that the properties of being black and being white are incompatible, such that no one thing can be both black and white. This reasoning is preserved here insofar as scorekeeping principles on sentences are generated from scorekeeping principles on frames. So we can say that the reason no coherent scorecard contains $\oplus\langle a \text{ is gray}\rangle$ and $\oplus\langle a \text{ is white}\rangle$ is that these are positions of the form $\oplus\langle x \text{ is gray}\rangle$ and $\oplus\langle x \text{ is white}\rangle$, and score is kept in accordance with the principle $\oplus_\alpha\langle x \text{ is gray}\rangle \vdash \ominus_\alpha\langle x \text{ is white}\rangle$

Now, in Chapter Two, we considered certain formal definitions of possible worlds, widely appealed to in laying out formal possible worlds semantic frameworks. We considered first the following definition:

> A possible world w is any function $f : \mathcal{A} \to \{true, false\}$.

We noted that, in order to ensure that the "worlds" provided by this definition were genuinely possible, we had to add the qualification that no subset of $\mathcal{A}$ whose members are jointly incompatible be mapped to *true*. However, given that extra-worldly semantics with explanatory ambitions was supposed to be giving an *account* of incompatibility in terms of possible worlds, this was problematic. We've now given an account of incompatibility in

terms of the pragmatic relation of preclusive consequence, where what it is for a sentence φ is incompatible with a sentence ψ is for *commitment* to φ to *preclude entitlement* to ψ, and vise versa. The relevant notion of an inconsistent set of sentences here—a set whose members are jointly incompatible—is a set S such that, for any scorecard σ, if $\oplus_{\alpha_1}\langle p\rangle \in \sigma$, for all $p \in S$, then there is a $q \in S$ such that $\ominus_{\alpha_1}\langle q\rangle \in \sigma$. That's just to say that this is a set of sentences such that one cannot be both committed and entitled to all of them, since commitment to all of the members of the set precludes entitlement to some. And it's clear that the notion of worlds, thus defined, corresponds, once again, to the notion of maximal, coherent, single-player scorecards. Such a scorecard σ determines a value for each atomic sentence p, *true* if $\oplus_{\alpha_1}\langle p\rangle \in \sigma$ and *false* if $\ominus_{\alpha_1}\langle p\rangle \in \sigma$, and it conforms to the coherence requirement since there is no sentence p such that $\oplus_{\alpha_1}\langle p\rangle \in \sigma$ and $\ominus_{\alpha_1}\langle p\rangle \in \sigma$.

We also considered, in Chapter Two, a somewhat more sophisticated definition of possible worlds that defines them as first-order models that conform to certain "meaning postulates." These meaning postulates are standardly specified in first-order or higher-order quantificational logic. Assuming the same three objects, a, b, and c, exist across all possible worlds, a possible world for our toy language can be understood simply in terms of a function that maps each basic 1-place predicate ("white," "gray," and "black") to a set of objects, and each 2-place predicate ("lighter than," "the same shade as," and "darker than") to a set of pairs of objects. Any such function defines a "world," but, in order to define the set of *possible* worlds, the function needs to conform to certain "meaning postulates," for instance, the following:

$$\forall x(\mathbf{gray}(x) \rightarrow \neg\mathbf{white}(x))$$

Laying down this postulate rules out any "world" w in which there is some object x, such that $x \in V(\mathbf{gray})(w)$ and $x \in V(\mathbf{white})(w)$. Now, as we've already seen, this universally quantified conditional can be understood as expresses the following material scorekeeping principle:

$$\oplus_{\alpha}\langle Gx\rangle \vdash \ominus_{\alpha}\langle Wx\rangle$$

And requiring coherence, given this scorekeeping principle, puts the very same constraint on scorecards that the above postulate puts on worlds, ruling out any scorecard in which

there is some term τ such $\oplus_{\alpha_1}\langle G\tau\rangle \in \sigma$ and $\oplus_{\alpha_1}\langle W\tau\rangle \in \sigma$. So, once again, the notion of a possible world, thus defined, can be understood as a transposition, into worldly vocabulary, of the notion of a maximal, coherent, single-player scorecard.

It should be clear how our reconstruction of worlds in terms of scorecards explains why, for instance, there is no world in which both "*a* is gray" and "*a* is white" are true; commitment to both sentences cannot show up on a single coherent scorecard, since commitment to one sentence precludes entitlement to the other. So, this account explains, in scorekeeping terms, the set-theoretic facts that, in the context of an extra-worldly semantics, are supposed to (in part) explain the fact that "*a* is gray" and "*a* is white" are incompatible. Moreover, note also that, since each maximal coherent scorecard determines a value, $\oplus$ or $\ominus$, for each atomic sentence p, then, given our logical rules, a value is determined for each logically complex sentence φ for each scorecard. From CO and our negation rules, we have $\oplus_\alpha\langle\varphi\rangle \vdash \ominus_\alpha\langle\neg\varphi\rangle$ and $\ominus_\alpha\langle\varphi\rangle \vdash \oplus\langle\neg\varphi\rangle$, and so, applying our logically extended scorekeeping principles to a maximal coherent single-player scorecard, we'll have $\oplus_{\alpha_1}\langle\neg\varphi\rangle \in \sigma$ just in case $\ominus_{\alpha_1}\langle\varphi\rangle \in \sigma$, and $\ominus_{\alpha_1}\langle\neg\varphi\rangle \in \sigma$ just in case $\oplus_{\alpha_1}\langle\varphi\rangle \in \sigma$, and so on. Likewise, from CO, RV, and our conjunction rules, we have $\oplus_\alpha\langle\varphi\rangle, \oplus_\alpha\langle\psi\rangle \vdash \oplus_\alpha\langle\varphi \wedge \psi\rangle$, we have $\ominus_\alpha\langle\varphi\rangle \vdash \ominus_\alpha\langle\varphi \wedge \psi\rangle$, and we have $\ominus_\alpha\langle\psi\rangle \vdash \ominus_\alpha\langle\varphi \wedge \psi\rangle$, and so $\oplus_{\alpha_1}\langle\varphi \wedge \psi\rangle \in \sigma$ just in case $\oplus_{\alpha_1}\langle\varphi\rangle \in \sigma$ and $\oplus_{\alpha_1}\langle\psi\rangle \in \sigma$, and $\ominus_{\alpha_1}\langle\varphi \wedge \psi\rangle \in \sigma$ just in case $\ominus_{\alpha_1}\langle\varphi\rangle \in \sigma$ or $\ominus_{\alpha_1}\langle\psi\rangle \in \sigma$. Dually for disjunction. Given this fact, we can see how this account is capable of explaining, in scorekeeping terms, the standard set-theoretic assignment of semantic values to logically complex sentences in a possible worlds semantics. Recall from Section 2.3, these assignments are the following:

$$[\![\neg\varphi]\!] = W - [\![\varphi]\!]$$
$$[\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \cap [\![\psi]\!]$$
$$[\![\varphi \vee \psi]\!] = [\![\varphi]\!] \cup [\![\psi]\!]$$

Consider, for instance, how our scorekeeping account of negation, coupled now with our account of possible worlds in terms of scorecards, explains the fact that the set of worlds assigned to $\neg\varphi$ will be the complement of the set of worlds assigned to φ. On our account of negation provided, if one is *precluded from being entitled* to φ, then one is *committed* to $\neg\varphi$. So, the set of (maximal, coherent, single-player) scorecards that contain $\oplus_{\alpha_1}\langle\neg\varphi\rangle$, will be

just those that contain $\ominus_{\alpha_1}\langle\varphi\rangle$, and, since each scorecard contains either $\oplus_{\alpha_1}\langle\varphi\rangle$ or $\ominus_{\alpha_1}\langle\varphi\rangle$, the set of scorecards that contain $\ominus_{\alpha_1}\langle\varphi\rangle$ will be the complement of the set that contains $\oplus_{\alpha_1}\langle\varphi\rangle$. Similar explanations can be straightforwardly provided for the other definitions.

Before considering the philosophical significance of this last point, we should note that not only can we reconstruct possible worlds, but we can reconstruct any kinds of worlds that one might want: impossible worlds, partial worlds, and so on. For impossible worlds, we simply drop the criterion that the scorecards be coherent, and, for partial worlds, we drop the criterion that they be maximal. The notions of complex normative positions, characterized by such scorecards, where one may be committed and precluded from being entitled to some sentence, and there, make perfect sense on this interpretation and are capable of conceptually grounding formal semantic theories for non-classical logics which require impossible worlds, partial worlds, or so on. Thus we can, for instance, provide a similar reconstruction, in scorekeeping terms, of Kripke's (1965) semantics for intuitionistic logic, framed in terms of partial worlds that stand in as non-trivial inclusion relations to one another, or Dunn/Restall (Dunn 1996, Restall 1999) incompatibility semantics for relevant negation, recently developed by Berto (2015) as a general account of negation, which involves the additional assumption that such worlds that stand compatibility and incompatibility relations, or Fine's (2017) truth-maker semantics, or what have you. These basic "worldly" notions that figure into all of these semantic theories can be reconstructed on this framework, and the functioning of the semantic theories themselves, based on these notions, can be explained.

5.8 Elucidatory and Explanatory Models, Revisited

At the end of Chapter One, I introduced a distinction between *elucidatory* models in semantics and *explanatory* models. Now, in thinking that an extra-worldly semantic theory falls on the latter side of this distinction, one might move from the fact that extra-worldly semantic models are *predictive* to the claim that they are *explanatory*. Consider, just to take the simple example that we considered in Chapter Two (Section 2.7), the fact that if φ entails ψ, then $\neg\psi$ entails $\neg\varphi$. This is a simple "prediction" of a possible worlds semantics,

given the definition of entailment, the semantics for negation, and some simple set theory. Recall, on a standard possible worlds semantics, φ entails ψ just in case $[\![\varphi]\!] \subseteq [\![\psi]\!]$, the semantics for negation tells us that $[\![\neg\varphi]\!] = W - [\![\varphi]\!]$, and it's a set-theoretic fact that complementation reverses the subset/superset relation: if $A \subseteq B$, then $C - A \supseteq C - B$. So, if $[\![\varphi]\!] \subseteq [\![\psi]\!]$, then $W - [\![\varphi]\!] \subseteq W - [\![\psi]\!]$, and so, if φ entails ψ then $\neg\psi$ entails $\neg\varphi$. This is a simple "prediction" of a possible worlds semantics, and, when we consider some concrete instances of it, it seems to be a good one. For instance, the theory predicts that, since "*a* is gray and *b* is white" entails "*a* is darker than *b*," it will also be the case that "It's not the case that *a* is darker than *b*" entails "It's not the case that (*a* is gray and *b* is white)," and, of course, that is indeed the case. A proponent of a possible worlds semantics might take this set of facts to provide an *explanation* of the fact that, if φ entails ψ, then $\neg\psi$ entails $\neg\varphi$. In Chapter Two, we argued that this was not so, and the reconstruction of possible worlds in scorekeeping terms that we have now provided shows why the set-theoretic structure that we have here is merely a *reflection* of this fact, not any explanation of it.

The transposition of these fact about possible worlds into this normative vocabulary is that if every (maximal, coherent, single-player) scorecard that contains $\oplus_{\alpha_1}\langle\varphi\rangle$ contains $\oplus_{\alpha_1}\langle\psi\rangle$, then every scorecard that contains $\oplus_{\alpha_1}\langle\neg\psi\rangle$ contains $\oplus_{\alpha_1}\langle\neg\varphi\rangle$. Why is this the case? Well, we *could* give the very same "explanation" as above. As we've just explained, the set of (maximal, coherent, single-player) scorecards containing $\oplus_{\alpha_1}\langle\neg\varphi\rangle$ will in fact be the complement of those containing $\oplus_{\alpha_1}\langle\varphi\rangle$, and so, from the fact that complementation reverses subset/superset relation, it follows that the set of (maximal, coherent, single-player) scorecards containing $\oplus_{\alpha_1}\langle\neg\psi\rangle$ will be a subset of those containing $\oplus_{\alpha_1}\langle\neg\varphi\rangle$. But is this really that explanation of the fact that, if φ entails ψ, then $\neg\psi$ entails $\neg\varphi$ that this scorekeeping framework is offering? Surely, it is not. First, on the framework here, the basic reason why it would be the case that the set of (maximal, coherent, single-player) scorecards containing commitent to φ also contain commitment to ψ would be that commitment to φ commits one to ψ, so any scorecard containing $\oplus\langle\varphi\rangle$, closed under this set of scorekeeping principles, will contain $\oplus\langle\psi\rangle$. Thus, this subset relation obtaining sets of (maximal, coherent, single-player) scorecards containing these commitments is not an *analysis* of an entailment relation obtaining between these sentences, but a *consequence* of

it, where this entailment relation is understood pragmatically in terms of a basic relation of committive consequence that a scorekeeper takes to obtain between these sentences. Now, given the account of negation that we've provided, according to which being *committed* to $\neg\varphi$ has the same discursive significance significance significance as being *precluded from being entitled* to φ, it's clear that the real explanation of the fact that, if φ entails ψ, then $\neg\psi$ entails $\neg\varphi$ essentially has to do with the following instance of *Reversal*:

$$\frac{\oplus\langle\varphi\rangle \vdash \oplus\langle\psi\rangle}{\ominus\langle\psi\rangle \vdash \ominus\langle\varphi\rangle}\ \text{RV}$$

This says that if *commitment* to φ *commits* one to ψ, then *preclusion of entitlement* to ψ *precludes one from being entitled* to φ. This basic fact about the normative structure of a discursive practice is one of key ingredients in the explanation of the fact that if φ entails ψ, then $\neg\psi$ entails $\neg\varphi$, and it doesn't even show up in the possible worlds "explanation" of this fact. Though the two semantic theories agree on their "predictions," the basic structure of explanations provided by the respective theories fundamentally differ.

Now, an extra-worldly theorist would presumably want to explain the instance of Reversal above in *pragmatic* terms. Following Stalnaker's (1978) approach to possible worlds pragmatics, we might propose that, when a speaker utters a sentence φ, the information state σ that characterizes what they're committed to and precluded from being entitled to, a particular set of possible worlds, is updated in the following way:[11]

Updates of states upon utterances: $\sigma[\varphi] = \sigma \cap [\![\varphi]\!]$

We can then say that, given their information state σ, a speaker is *committed* to a sentence φ just in case a speaker's information state *includes* the information expressed by φ. So,

An agent in state σ is *committed* to φ just in case $\sigma \cap [\![\varphi]\!] = \sigma$

Alternately, a speaker is *precluded from being entitled* to φ just in case that information state *excludes* the information expressed by φ. So,

[11] Unlike Stalnaker's proposal, the context states that we are treating as updated in this way characterize the commitments of *individuals* rather than the shared commitments of a *group*.

An agent in state σ is *precluded from being entitled* to φ just in case $\sigma \cap [\![\varphi]\!] = \varnothing$

We thus purport to explain the pragmatic fact that uttering "*a* is gray" and "*b* is white" commits one to "*a* is darker than *b*" and precludes one from being entitled to "*a* is lighter than *b*"in terms of the underlying possible worlds semantics. Likewise, we purport to explain the instance of Reversal specified above, which, in this context, becomes the principle that, for all sentences φ and ψ and states σ, if $\sigma \cap [\![\varphi]\!] = \sigma \Rightarrow \sigma \cap [\![\psi]\!] = \sigma$, then $\sigma \cap [\![\psi]\!] = \varnothing \Rightarrow \sigma \cap [\![\varphi]\!] = \varnothing$. The antecedent of this conditional will hold (for all sentences and states) just in case $[\![\varphi]\!] \subseteq [\![\psi]\!]$, for only then will it be the case that, if intersecting with $[\![\varphi]\!]$ doesn't remove worlds for σ, then neither will intersecting with $[\![\psi]\!]$, but if $[\![\varphi]\!] \subseteq [\![\psi]\!]$, then the consequent will hold as well, since if intersecting with $[\![\psi]\!]$ removes all the worlds, then so will intersecting with $[\![\psi]\!]$. In this way, the pragmatic relations, and, moreover, the relations between the pragmatic relations, are all understood as supervening on the semantic relations.

Perhaps, at the beginning of this dissertation, one would be inclined to think we have simply reached a certain kind of stalemate here, where the basic notions of one theory can be explained in terms of the basic notions of the other. However, as I have argued throughout this dissertation, these two orders of explanation are not on equal footing. Though we can genuinely explain the worldly contents appealed to in a possible worlds semantics by starting with a normative pragmatic account of discursive roles, as I have just shown in this chapter, the explanation cannot go the other way. The order of explanation proposed by the possible world semanticist essentially appeals to worldly knowledge in explaining speakers' knowledge of meaning, and this worldly knowledge, as I've argued, can be understood only as a reflection of semantic knowledge. Thus, while a possible world semantics may be deployed to *elucidate* the structure of semantic competence, it cannot be deployed to *explain* this semantic competence. I have argued, in Chapter Four, that semantic competence is to be understood in normative terms, and I have thus shown here that the worldly contents appealed to in the context of a possible world semantics—or any worldly semantic theory, for that matter—are to be understood as *products* of semantic competence, reifications of discursive norms.

5.9 Getting Real

Let me close this chapter by responding the most glaring objection to the theory of meaning that I have laid out, which will require moving beyond our simple toy language, which has now served its purpose. Our simple toy language, recall, has the predicates "black," "gray," "white," "lighter than," "darker than," and "the same shade as." In Chapter One, I said that, with these predicates, the "speakers" of our toy language can *say* that something's black, that something's gray, that something's white, and that something lighter than, darker than, or the same shade as something else. This, it may now be clear, is very much an objectionable thing to say, and a line of objection might go as follows:

> OBJECTOR: Can they *actually* say these things, though? Surely, they can't *really* say that something's *black*, can they?
>
> ME: Well, that depends: what do you mean by "black"?
>
> OBJECTOR: I mean what *you and I* mean by "black" when we look at my shoes, for instance, and say that they're black.
>
> ME: Oh, no, they certainly don't mean what *we* mean when we say that something's "black."
>
> OBJECTOR: Well what *do* they mean, then?
>
> ME: They mean what *they* mean, of course!
>
> OBJECTOR: But *what do they mean!?!*
>
> ME: I just spent two chapters telling you!
>
> OBJECTOR: But you *haven't* told me! You've defined the meanings of "black," "gray," "white," "lighter than," "darker than," and "the same shade as," all in relation to one another, and you've spoken of the speakers as grasping "the property of being black" in virtue of grasping these relations. However, for all that you've actually given the "speakers" of this toy language, the property expressed by the predicate "black" of their language could just as well be *the property of being red,* since the predicates "red," "pink," and "white" stand in the very same set of relations to one another that "black," "gray," and "white" do.
>
> ME: Clearly the correct thing to say here is not that the property grasped by speakers of the toy language, expressed by their predicate "black," could be

either the property of being black or the property of being red—that the linguistic rules somehow underdertmine the property expressed by the predicate. Rather, as I've already indicated, the predicate "black" of the toy language expresses *neither* of these properties. After all, part of what it is to grasp that something is black, in our sense of the term "black," is to grasp that, if something's black, then it can't be colored, and the speakers of the toy language have no scorekeeping principle they would express with this conditional. Likewise, for red, one must know such things as that if something's red, then it *is* colored. So, the speakers of the toy language grasp neither of these properties. But *these* properties are the worldly correspondents of the rules governing the use of *our* expressions "black" and "red," whereas "the property of being black" that I've officially defined is the the worldly correspondent of *their* expression "black," and *this* property is perfectly determined by the rules governing the use of their expression "black," because this property *just is* a reification of those rules.

OBJECTOR: But that's not the property of being black! Indeed, that's not any real property at all! Perhaps it's a *toy* property that corresponds to an expression of this *toy* language, but I don't want to know about *toy* properties; I want to know about *real* properties! That's what you've promised us an account of, and so far you haven't given us one!

ME: Well, accounting for the properties of this toy language was supposed to provide a simple model for accounting for the properties that are the worldly correspondents of the expressions of our languages, but I suppose I should now say—or, better, show—how this model can be put to use.

So, let's get real. Switching up the example to consider what is for some reason the philosopher's favorite color, let's consider the meaning of the English predicate "red," which expresses the property of being red, a real property on which we speakers of a real language have a grip. On the account I've given, this property is to a reification of the norms governing the use of "red." Let us consider these norms.

First, this predicate belongs to a family of other color predicates, and its use stands in normative relations to their use. These sorts of intra-family relations are the ones that we've explicitly considered here, and so we now have a very clear sense of how to understand them. For instance, commitment to a sentence of the form "x is red" commits one to "x is colored," precludes one from being entitled to "x is green," commits one, along with "y is pink," to "x is darker than y," is a consequence of commitment to "x is crimson," and so on. The structure of *value* or *brightness*, explicated above in terms of the norms governing the use of the predicates "darker than," "lighter than," and "the

same shade as," can be understood as constituting a fragment of the actual structure here, though there are other, orthogonal dimensions we now must consider, which can be represented as additional dimensions in a color space. Particularly, there are the additional relations of *hue*, concerning *what* color something is, and *saturation*, concerning *how* colored something is. Considering just the former dimension, there is a relation of "closeness" in hue, according to which we can say that the property of red is closer to orange and violet than it is to blue, yet closer to blue than it is to green. Clearly, there is quite a bit more structure here than that considered for our toy language, but this structure can be articulated in just the same way.

However, a grasp of the meaning of "*red*" requires much more than a grasp of these intra-family relations. In order for really account for the conceptual significance of the specific location in the three dimensional color space that is to be identified with the color *red*, we must consider at least some scorekeeping principles relating the use of "red" to the use of other discursively significant non-color expressions. For one, it's clearly crucial to the meaning of red that, if one's looking at something red in good lighting, one can see and thereby know that it's red. There are other connections, however, that, while not each essential individually, could not be wholly removed with "red" retaining its basic conceptual significance. For instance, the color red often has a sense of warning, and this can be understood, in part, in terms of the fact that stop signs are red, and these tell one to stop, as do the red traffic lights. We might also note that redness is associated with ripeness, as ripe tomatoes, ripe raspberries, and ripe strawberries are red (though red blackberries are unripe), and, if something's ripe, it's good to eat. We could go on to add many other connections, but we might stop there. All of these statements are to be understood, on this account, as expressions of scorekeeping principles such as that commitment to "x is a stop sign" commits one to "x is red," commitment to "x is a tomato" along with "x is ripe" commits one to "x is red," and so on. In this way, by connecting the use of the predicate "red" with other practically significant expressions, the discursive significance of "red" is more than a point in an abstract structure.

Now, it might seem that, in order to truly appreciate the discursive roles of practically significant expressions such as "sees," "stops," and "eats," we need to broaden the con-

ception of MOVE to beyond mere assertions to include, for instance, such things as acts of seeing, and perhaps even acts of stopping, and acts of eating.[12] This would be a radical modification of the theory, and, if such a modification is necessary, it's hard to see how anything like the simple model provided in connection to the toy language could be adequate. However, we can integrate the varied practical significance of these expressions all within the basic framework of discursive role semantics, without radically modifying it so that MOVE includes anything other than acts of uttering declarative sentences. The key thought, which I spell out in detail elsewhere (Simonelli, M.S.e), is that this is possible as long as the language includes sentences that involve the *attribution* of such acts. Consider just acts of seeing. As we've said, the meaning of "red" is essentially bound up with the fact that if one's looking at something red in good lighting, one can see and thereby know that it's red. Crucially, we need not radically change the basic shape of the semantic theory in order to accommodate this fact, for we can simply add scorekeeping principles such as that commitment to "x is red," "n is looking at x," "n has color vision," and "The lighting is good," commits one to "n sees that x is red." Insofar as we acknowledge that seeing is a way of being entitled, we can introduce the notion of a sensible quality as the worldly correspondent of a predicate whose application is one to which one can be entitled through perceptual attribution. Spelling out the details of such an account is a project left for other work, but an account of this sort can be relatively straightforwardly accommodated in the framework put forward here. Similarly for other non-assertive acts to which the use of "red" is related, such as stopping at stop signs and eating of ripe tomatoes.

Though this is no more than a gesture at a full account, it should suffice to show that we can, at least in principle, give a perfectly adequate account of the property of being red, grasped by speakers of English, in just the sort of framework developed here, as the reification of scorekeeping principles such as the following:

$\oplus_\alpha\langle x \text{ is red}\rangle \vdash \oplus_\alpha\langle x \text{ is colored}\rangle$

$\oplus_\alpha\langle x \text{ is red}\rangle \vdash \ominus_\alpha\langle x \text{ is green}\rangle$

$\oplus_\alpha\langle x \text{ is red}\rangle, \oplus_\alpha\langle y \text{ is orange}\rangle, \oplus_\alpha\langle z \text{ is blue}\rangle \vdash \ominus_\alpha\langle x \text{ is closer in hue to } y \text{ than } z\rangle$

[12] Consider the proposal of Kukla and Lance (2009), drawing on Belnap (1990).

$\oplus_\alpha\langle x \text{ is a stop sign}\rangle \vdash \oplus_\alpha\langle x \text{ is red}\rangle$

$\oplus_\alpha\langle x \text{ is a tomato}\rangle, \oplus_\alpha\langle x \text{ is ripe}\rangle \vdash \oplus_\alpha\langle x \text{ is red}\rangle$

$\oplus_\alpha\langle x \text{ is red}\rangle, \oplus_\alpha\langle \text{The lighting is good}\rangle, \oplus_\alpha\langle x \text{ is in } n\text{'s line of sight}\rangle, \oplus_\alpha\langle n \text{ has color vision}\rangle \vdash \oplus_\alpha\langle n \text{ sees that } x \text{ is red}\rangle$

And so on . . .

Of course, the full set of scorekeeping principles that would have to be specified in order to give a complete definition of the property of being red, defining it as a reification of these principles, would be massive. Though it will be finite at any given point in time, since any actual language only contains a finite number of atomic sentences, it will be astronomically large, and, moreover, will always be growing and otherwise changing, since language is a dynamic, evolving entity. So there's no reason to aspire to such a specification. Still, certain key clusters of scorekeeping principles, for instance, those constitutive of the fact that something's being red is its instantiating a sensible quality, are going to be worth spelling out in detail. This is, as I've already explicated, is a task for lexical semantics, understood as the systematic articulation of material scorekeeping principles. But lexical semantics, done here, just is the metaphysics of perception!

Consider, for instance, a recent dispute between James Conant (2020) and John McDowell (1994, 2002) on the proper articulation of the notion of *seeing*, conceived of as an actualization of a perceptual capacity.[13] Conant takes issue with McDowell's conception of seeing, according to which an act of seeing *entitles* one to judgments, putting one *in position* to know, but is somehow less committal than judgment and so does not itself constitute an act of knowing. In reading Conant's arguments, one might be inclined to wonder about the nature of the dispute. On the one hand, it seems to be a metaphysical dispute about the form of the capacity for perceptual knowledge, not a dispute about the English expression "sees," but, on the other hand, Conant's arguments appeal to our intuitions, as competent speakers of English, about the relative priority of different uses of this expression. The account of metaphysics as reified lexical semantics proposed here makes clear sense of Conant's methodology, for the metaphysical structure at issue here

[13]See also Stroud (2018a) for a development of this line of criticism against McDowell.

can be understood as a reification of scorekeeping principles involving the term "sees."[14] If we agree with Conant, then, transposing at least one key aspect of his account into the current vocabulary, we'll have that if one is committed to a claim of the form "*n* sees that *x* is red," then not only is one committed to "*n* is entitled to ⟨*x* is red⟩" and committed to "*x* is red" oneself, but also committed to "*n* is committed to ⟨*x* is red⟩." Things are further complicated when we bring in the scorekeeping principles constitutive of the notion of something's *merely looking* red, articulating the relationship between "sees," "is," and "(merely) looks." The core idea of the Sellarsian (1956) account the according to which what one is doing in undertaking a commitment to a claim of the form "*x* looks red" is holding back the commitment to the claim "*x* is red" that one undertakes in undertaking a commitment to "I see that *x* is red." Once again, systematically spelling out such an account in terms of scorekeeping principles is beyond the scope of the current project; the point is just to be clear that such an account has a place in the context of the semantic theory.

Finally, let us be clear that, though we've been speaking of English expressions here (since that's the language in which this dissertation is written), the property of being red surely can't be defined as a reification of the *English* predicate "red." Clearly, speakers of Spanish, German, Japanese, and various other natural languages are capable of saying of things that they're red, and, moreover, it is certainly possible that the English language could have never developed at all with speakers of other languages still having a grip on the property of being red and being able to ascribe it to things. Now, on a worldly semantic theory, one provides what Sellars (1956) calls a "relational" account of meaning, according to which sameness of meaning across different languages is in terms of different words of different languages all being related to the same extra-linguistic entity, be it an object, property, or relation. Consider, the following sentence:

1. The word "rojo," in Spanish, means *red*.

On a relational analysis of "means," when we say that the word "rojo," in Spanish, means *red*, what we're doing is relating the Spanish word "rojo," picked out with the phrase "The

[14]For a fuller and more subtle account of the normative/modal correspondence, as it applies here to specifically concern the vocabulary of *agentive* modality (talk of "capacities"), see Simonelli (2020).

word 'rojo"' with a particular non-linguistic entity, the property of being red, picked out by a special referential (rather than predicative) use of the word "red," achieved here by italicization. On the account offered here, which is owed to Sellars (1956, 1979), rather than functioning to pick out some non-linguistic thing, the italicized "red" is playing a special sort of predicative role, functioning to characterize the word "rojo" as playing the same role, in its home language, as it itself plays in its home language. Now, of course, there will never be *complete* overlap in role between words of different languages, but will have to be *substantial* overlap for a sentence like (1) to be assertable, and, clearly, in the case of "red" and "rojo," there is.[15]

Though the property of being red corresponds to the English predicate "red," the Spanish predicate "rojo," and the German predicate "rot," not all properties expressible by simple predicates of one language will correspond to those in other languages. Consider, for instance, the property of being a chair.[16] As speakers of English, we grasp this property. We grasp, for instance, that, if something's a chair, then it's something one can sit on, that it generally has a back (but not always, as in the case of bean bag chairs), that it might be hard, like a kitchen chair, or cushiony, like an armchair or recliner, and so on. This property, however, simply doesn't belong to the network of simple properties on which a native German speaker comes to have a grip through learning German. In German, there's the predicate "Stuhl," which can be correctly applied to kitchen chairs, but not armchairs and recliners, there's the predicate "Sessel," which can be correctly applied to armchairs and recliners, but not kitchen chairs, and there's "Sitz," which can be applied to anything on which one might comfortably sit, but that applies to tree stumps and comfortable rocks no less than it applies to chairs—there's no one simple predicate that can be applied to all and only *chairs*. As such, the property of being a chair simply doesn't belong to the network of simple properties on which a native German speaker comes to have a grip through learning German.[17] The lack of correspondence between the properties grasped

[15]Importantly, as Lance and Hawthorne (1997) have argued at length, (1) should really not be construed as a *descriptive* claim at all, but, rather, an *normative* claim. So, an English speaker who is committed to (1) will take Spanish speakers to be bound, in using the word "rojo," to the norms that they take to govern the use of the word "red," whether or not Spanish speakers generally acknowledge all of these norms themselves.

[16]Thanks to Jim Conant for this example.

[17]Of course, that's not to say that the property of being a chair can't be *introduced*; a German speaker

by the English speaker and the properties grasped by the German speaker is understood in terms of the lack of correspondence between the rules governing the use of the English and German predicates; while many English words correspond sufficiently closely in role to a German word, such that speakers can be said to have a grip on the same properties, "chair" does not. All of this makes perfect sense on the account offered here according to which properties are reifications of discursive roles.

5.10 Conclusion

In this chapter, I have given an account of quantified modalized conditionals such as, "If something's black, then, necessarily, it's darker than anything white" as functioning to express the scorekeeping principles that determine the semantic significance of basic expressions such as "black" and "darker than." With the use of this formal framework, I have shown how we can think of the worldly contents appealed to in worldly semantic theories as reifications of these scorekeeping principles. This provides a satisfying metaphysical and epistemological story of the metaphysical entities appealed to in the context of worldly semantic theories and our grasp of them. It also vindicates our claim that such theories get things explanatorily backwards. Rather than our grasp of the rules governing the use of linguistic expressions asymmetrically depending on our grasp of such things as properties and relations, our grasp of properties and relations is really nothing other than our grasp of the rules governing the use of linguistic expression, transposed into a worldly mode. This account makes room for an account of the real relation between language and the independent world, and that is where we now turn.

can surely learn what an English speaker means when they use the word "chair," and that's what one does when one learns English. But this is explicated in terms of the properties that one already grasps through learning German.

6

The Language-World Relation

In the previous chapters, I developed an account of meaning according to which the "world," as it is appealed to in semantic theorizing, cannot be assumed to be anything more than a reification of the rules governing the use of linguistic expressions. This raises a question about the *world*, apart from our language, and how language relates to it. Now, if one is inclined to a certain sort of linguistic idealism, one might think that the "world" we've been discussing this whole time just is the world. Linguistic idealism, however, is not the intended upshot of this dissertation. The intended upshot, rather, is a sort of critical realism, one that avoids the Myth of the Given to which worldly semantic theories fall prey but which also avoids the form of linguistic idealism that plagues certain followers of Sellars, and most followers of Brandom. It is time to answer Stanley's (2006) challenge, showing that, far from being "useless for the philosophical project of understanding the language-world relation," discursive role semantics, unlike worldly semantics, is able to provide us with an understanding of the complex and multi-faceted relation between language and the world.

6.1 Beyond Saying and Doing

Let us start with an characterization of the expressivist logic we provided in the previous chapter, using the toolkit Brandom (2008) develops in *Between Saying and Doing*. Brandom's main concern in that book is what he calls "pragmatically mediated semantic relations." These are relationships betwen *vocabularies* that can be understood only through consideration of the underlying *practices* in which the meanings of the expressions belonging to these vocabularies and others are conferred by their use. The core

tool Brandom develops for representing these pragmatically mediated semantic relations is what he calls "meaning use diagrams." Consider the following diagram, which characterizes the sense in which the vocabulary of modalized conditionals, developed in the previous chapter, can be understood as *elaborated from* (L) and *explicative of* (X) a fact-stating discourse:

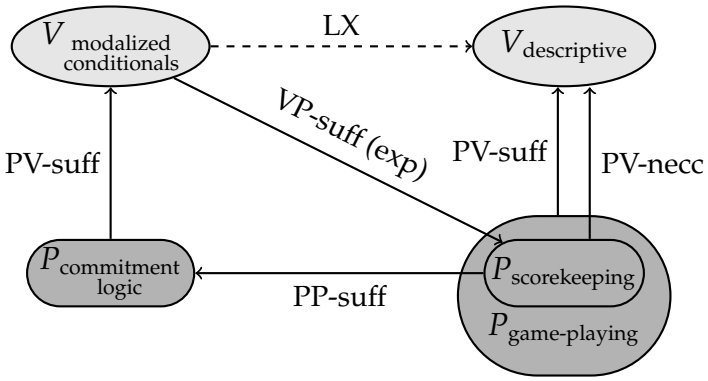


Figure 6.1: LX-ness of Modalized Conditionals

The arrows here represent various sorts of *sufficiency* and *necessity* relations between practices and vocabularies. The practice of keeping score is PP(practice-practice)-sufficient for the practice of deploying the commitment logic, in the sense that, if one is capable of engaging in the first practice, one's abilities can be "elaborated" into those required for engaging in the second practice, hypothetically attributing commitments to arbitrary speakers, attributing consequential commitments through application of one's scorekeeping principles, and affirming conditionals accordingly. This practice, spelled out with a Fitch-style proof system, is PV(practice-vocabulary)-sufficient for the vocabulary of modalized conditionals. This vocabulary, in turn, is VP-sufficient to specify the scorekeeping practice which is an essential component of any fact-stating discourse, though, as the parenthetical indicates, this VP-sufficiency comes through the sufficiency of the vocabulary to *express* scorekeeping principles rather than to explicitly *state* them, as one would with the normative vocabulary we've deployed as a metavocabulary here.

This last point is important, for, as we've emphasized here, the modalized conditionals we've considered *express* but are not *about* scorekeeping principles. For instance, the conditional "If something's copper, then, necessarily, it melts at 1085° C" *expresses* the

principle of scoring anyone committed to a sentence of the form "*x* is copper" to be committed to "*x* is melts at 1085° C," but it's not *about* this principle or the expressions it concerns. Rather, insofar as it's about anything, it's about *copper* and chemical properties; it says of *copper*, a particular chemical substance, that it melts at 1085 degrees Celsius. So, though this vocabulary of modalized conditionals stands in vocabulary to practice relation of expressing scorekeeping principles, it seems that it also stands in a vocabulary to *world* relation—stating *facts*, for instance, about the exceptionless propensities of things in the world such as copper. And so we may well wonder not just about the relation between this vocabulary and the *practice* that it is elaborated from and explicative of, but of the relation between it and the *world*, which contains, among other things, copper which behaves in certain ways under certain conditions.

This question about the relationship between language and the world might seem particularly pressing given the account of properties that I have articulated here, according to which such things are understood, at least in the first instance, as reifications of discursive roles. Copper and its chemical properties, however, clearly seem to be *independent* of our linguistic practices and vocabularies. And not only that, but it seems that the fact that copper has the chemical properties that it does must, in some way, *account* for the fact that we have the scorekeeping principles concerning "copper" that we do. The reason why someone who's informed about the chemical properties of copper scores someone who's committed to "*x* is copper" as thereby committed to "*x* melts at 1085° C," is because *copper melts at 1085° C* and the scorekeeping principles one has concerning "copper" are responsive to the facts concerning copper. So, what is the relationship between these practices and the vocabularies conferred by them, on the one hand, and the independent world that they concern, on the other? In diagrammatic form, we might put this question as follows:

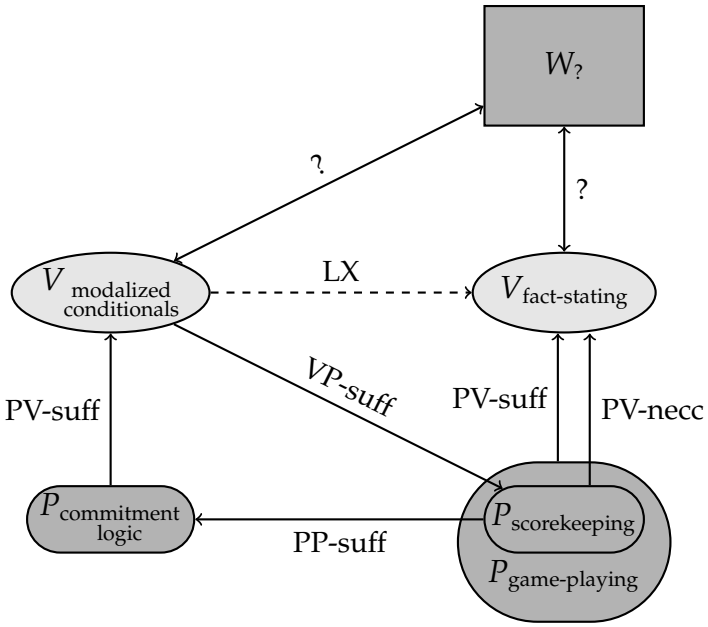


Figure 6.2: The Language-World Relation?

Extending meaning use diagrams to include sharp rectangles, meant to represent bits of the world, we can use them to consider the relations between our vocabularies, fact-stating and modal, and the bits of the world that we presumably want to think of these vocabularies as describing. In early work, Brandom explicitly avoids any talk of such relations, aiming to reconstruct seemingly "word-world" relations such as reference entirely in terms of "word-word" relations. However, in recent and unpublished work, Brandom (2019a, 2019b) has taken up the task of articulating the relationship between descriptive or representational vocabularies, on the one hand, and the world they describe or represent, on the other. Though, as we will see, we will have to move beyond Brandom's account, it will provide a good starting point for our investigation.

Let us begin by recalling the distinction between two *flavors* of entailment and incompatibility relations. The sorts of entailment and incompatibility relations with which we've concerned ourselves in our formulation of discursive role semantics and our commitment logic are *normative* ones, understood in terms of normative relations of consequential commitment and preclusion of entitlement between acts of assertorically using sentences or predicates. Commitment to one claim can *commit* one to some claim or preclude one from being *entitled* to another claim. In our articulation of properties and possible worlds,

however, we've also considered entailment and incompatibility relations of an *alethic* flavor, where these relations are understood as the alethic relations of necessitation and preclusion of possibility between properties or states of affairs. The obtaining of one state of affairs can *necessitate* the obtaining of another state of affairs or preclude another state of affairs from being *possible*. Now, we might ask, what is it to take oneself, in thinking and speaking, to be representing properties that belong to the *objective world*, which is independent of one's subjective acts of thinking and speaking? Brandom's basic answer to this question is that what it is to represent things as instantiating objective properties, relations, and states of affairs—constituents of the independent world—is to take oneself to be *normatively* bound, in one's assertoric use of predicates and sentences, by the *alethic* entailment and incompatibility relations that obtain between the properties, relations, and states of affairs one is representing. This regulative relation between these worldly entities, articulated in terms of alethic modal relations, and our linguistic activity, articulated in terms of normative relations, is one that Brandom (2019a) calls *semantic governance*. The worldly states of affairs, properties, and relations we represent, if we are to be counted as representing them, must *govern* or *reign over* our linguistic acts in the sense that the normative standards that we hold ourselves to in performing these acts must be taken to be inherited from the alethic structure of those states of affairs, properties, and relations themselves.

The notion of "governance" Brandom appeals to here has been explicated at length by John Haugeland (1998) in terms of what he calls "*beholdenness*" to objects. To consider this notion, as Haugeland spells it out, let us first note that the things in the world that we take to govern our vocabularies are not *bare* objects, but essentially things of certain *kinds*. Insofar as they are the things that they are, they are the kinds of things that they are, and that means that they necessarily do certain things and can't possibly do certain other things. That is to say, they conform to certain *constitutive standards*. A piece of copper, for instance, essentially behaves in a certain way, melting at 1085° C, conducting electricity, falling to the earth when dropped in solid form, and so on. Something's doing these things, behaving in all the ways that copper does, is constitutive of its being copper. This is in fact a consequence of the account of properties we have offered, according to which properties

are constituted by the metaphysical structure articulated with modalized conditionals such as "If something's copper, then, necessarily, it melts at 1085° C." Where the relevant properties are *kinds*, which specify what an object *is*, rather than merely something an object *has*, the modalized conditionals that articulate the contents of the relevant kind terms articulate the standards that different things in the world conform to insofar as they are what they are. A given piece of copper, for instance, cannot be what it is and not do what copper does. In this sense, constitutive standards cannot possibly be violated. However, insofar as we conceive of things in the world as genuinely *independent* of our conceptualization of them, as being what they are independently of what we take them to be, we must nevertheless conceive of things as potentially violating the standards that would constitute their being what we take them to be, calling upon us to reconceptualize them as differently constituted.[1]

In spelling this out, we may distinguish between two kinds of reconceptualization: what I'll call "mundane reconceptualization" and "constitutive reconceptualization."[2] Mundane reconceptualization takes place at the level of the particular things. In such cases, we simply realize that some things are not what we took them to be. For instance, if we place a piece of copper in water and it dissolves, our reaction is not to think that a piece of copper has violated the standards constitutive of what it is to be copper, but, rather, that this thing isn't actually copper or perhaps that the liquid we placed it in is not actually water. On Haugeland's account, existential commitment—commitment to the existence of copper, orca whales, ribozymes, black holes, the Higgs field, or what have you—is a sort of resiliency in the face of apparent violations of constitutive standards. In cases of apparent violations, rather than thinking that the constitutive standards have really been violated, we consider first whether we really have a sample of that kind, we double check our instruments, we see what could have gone wrong in the experiment, and so on. Insofar as we are existentially committed to things, we seriously consider the possibility

[1] As Kukla and Lance (2014) have emphasized, largely in response to Haugeland, talk of objects "governing" or having "authority" over us is a metaphor that ultimately needs to be spelled out. Really, the only things doing the "holding," as it were, are other discursive practitioners who hold one another to the standard of being responsive to the objects.

[2] This terminology is drawn from Haugeland's (1998) distinction between "mundane capacities" and "constituitive capacities."

of the violation of constitutive standards only as a last resort. However, holding onto this possibility as a genuine one is necessary in order to conceive of the things to which we are existentially committed as being what they are independently of we take them to be. This possibility's obtaining, however, is not a possibility defined within a pre-existing space of possibilities, for, if some class of objects to which we are existentially committed systematically violate their constitutive standards, this undermines their very status as being those objects, since, if the constitutive standards are violated, then *those* objects, the ones constituted by those constitutive standards, *aren't*.[3] This sort of self-undermining brings forth a second sort of conceptual revision, constitutive reconceptualization, where we have revisions of our conception of the constitution of reality, revising our conception of its basic alethic structure.

This is the conceptual engine at the core of scientific practice. Scientific theorizing is, in large part, the articulation of constitutive standards of observed objects or objects theoretically postulated to explain the behavior observed objects, and, in scientific practice, we are constantly giving these articulated objects chances to violate their constitutive standards. The fact that objects articulated by scientific theories *don't* violate their constitutive standards, when they or their effects are observed under a wide range of circumstances, is reason to think that objects constituted by those standards really exist in the world independent of our linguistic practices.[4] The most groundbreaking revolutions in scientific theorizing occur when objects appear to conform to certain constitutive standards articulated by a scientific theory under a very wide range of circumstances, but are then found to violate those standards in certain specific circumstances. For instance, in terms of repeated observations of objects conforming to the constitutive standards articulated by the theory, the Newtonian theory of material bodies and the gravitational forces they exert on one another was among the most successful scientific theories in history. It turns out, however, that, though Newton's theory makes very accurate predictions of the observable behavior of material bodies across a very wide range of circumstances, material bodies

[3] I am putting things slightly less paradoxically than Haugeland (1998), who speaks of this idea in terms of what he call "the excluded zone," which he describes as "a non-zero extension of the conceivable beyond the possible—that is *in fact* empty," (333).

[4] This is an expression of the so-called "no miracles" argument for scientific realism (Putnam, 1975a).

do not actually conform to the constitutive standards it articulates. This was revealed most strikingly in Mercury's failure to revolve around the Sun as a Newtonian body must. Upon this observation of the violation of constitutive standards, the scientific community, committed to the existence of Newtonian bodies, first made various attempts at mundane reconceptualization, aiming to maintain that the violation of constitutive standards apparently revealed by behavior of Mercury was merely apparent. For instance, another planet, dubbed "Vulcan," was postulated whose gravitation force would explain the irregularity in Mercury's orbit. None of these attempts at mundane reconceptualization proved successful, and constitutive reconceptualization eventually came in the form of Einstein's theory of gravity in terms of the curvature of spacetime which correctly predicated the orbit through a radically distinct conception of the phenomenon at hand. Material bodies, understood in the context of Einstein's theory of relativity as essentially such as to curve spacetime, are fundamentally different kinds of things—bound by different constitutive standards—than Newtonian bodies.

Of course, not all instances of constitutive reconceptualization are as dramatic as the transition from Newtonian mechanics to Einsteinian mechanics, which radically transformed our conception of the basic structure of physical reality. Myriad more local instances of constitutive conceptualization and reconceptualization have taken place through the course of the history of natural scientific inquiry, yielding the resilient body of scientific understanding we have today. All of this, I take it, is an unpacking of the notion of "semantic governance," of letting the meaning-constitutive norms of the practice be determined by the of objects of theoretical inquiry, of being beholden to the objects. Being beholden to the objects requires responsiveness, in the construction of scorekeeping principles, to what they what they do, revising the scorekeeping principles constitutive of our conception of what objects are if they do things that they constitutively cannot, given such a conception. Because of this engine of conceptual revision that obtains in virtue of semantic governance by genuinely independent objects, a *counterfactual* relation obtains between linguistic practices that have this structure and the objects that govern those practices: if the objects had been different, the norms governing practice would be different. As Brandom (2019b) articulates it, this is, complementary to the *normative*

relation of *semantic governance*, an *alethic modal* relation of *epistemic tracking*, which has the opposite directionality, going from the linguistic practice to the world. So, the linguistic practice is *semantically governed by* the objects, properties, and relations in the world, and the objects, properties, and relations in the world are *epistemically tracked by* the linguistic practice. Here is what we might call a "language-world" diagram that Brandom (2019b) gives us, depicting these two relations:

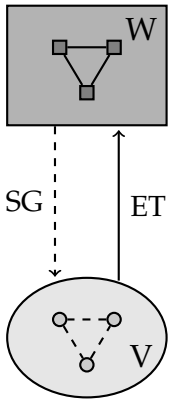


Figure 6.3: SG and ET w.r.t. Fact-Stating Vocab

Here, the dotted lines depict normative relations and the solid lines depict alethic modal relations. The big ellipse is a vocabulary, a bit of language, the little ellipses within it are claims that can be made with the use of that vocabulary, and the dotted lines going between them are normative relations of entailment and incompatibility that obtain between acts of making these claims. The big rectangle is the bit of the world that the vocabulary is both about and responsive to, the little rectangles are within it are states of affairs that may or may not obtain in the world, and the solid lines going between them are alethic modal relations of entailment and incompatibility that obtain between these states of affairs. The two arrows show that the bit of the world depicted by the big rectangle stands in the normative relation of *semantically governing* the corresponding bit of language, and the bit of language depicted by the big ellipse stands in the alethic modal relation of *epistemically tracking* the corresponding bit of the world.

If both of these relations obtain, then having a grip on the norms governing the use of linguistic expressions, and conceiving of this grip in worldly terms, just is to have a

grip on the alethic structure of reality. We can put this in terms of the following diagram (Brandom 2019b):

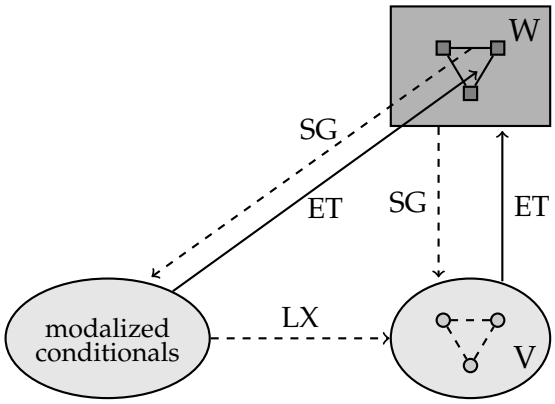


Figure 6.4: SG and ET w.r.t. LX Modalized Conditionals

Brandom says that the diagonal ET and SG relations shown here, obtaining between modal *vocabulary* and modal *relations* constitutive of the structure of reality, are *induced by* and *deducible from*, the vertical ET and SG relations and the horizontal LX relation elaborated above. Though he never explicitly says how this induction and deduction go, it's not hard to fill in the details. Insofar as the base vocabulary is semantically governed by and epistemically tracks features of the world, the alethic structure of the world corresponds to the normative structure of the base vocabulary, and so the modalized conditionals that express the normative structure of the discursive practice will articulate the alethic structure of the world. So, when the fact-stating vocabulary of which modal vocabulary is LX is semantically governed by and epistemically tracks objective properties and relations, this modal vocabulary can be understood as articulating the structure of objective reality. This diagram thus shows that the account of properties of alethic reifications of the norms governing is compatible with saying that some properties, at lesat, are more than mere reifications, for the very alethic modal structure that is constitutive of these properties may well be instantiated in the world. In such a case, the modal vocabulary that articulates that structure, expressing the scorekeeping principles of a vocabulary that is governed by and tracks features of the world, can be understood as describing the structure of reality.

At this point, it is worth contrasting the version of modal normativism put forward

here, with this additional aspect of the account on the table, with Thomasson's (2020) version of modal normativism, discussed in the previous chapter. There are two key points of contrast. The first is that Thomasson's account of modal normativism is explicitly restricted to specifically *metaphysical* modality, whereas the account here applies just as well to *nomological* modality as it does for metaphysical modality. Indeed, as we'll see in detail below, there is rarely a clean-cut way of delineating the two modalities. For instance, insofar as metaphysical modal truths are just those claims that express semantic rules, grasped by semantically competent speakers, it seems clear that "Whales are animals" must be a metaphysical necessity; clearly, someone who doesn't know that a whale is an kind of animal doesn't know what "whale" means. But then what should we say about "Whales are mammals," which is the product of empirical inquiry? It seems odd to say that there is a fundamentally different kind of fact expressed by these sentences, a "semantic" one expressed by the first and an "empirical" one expressed by the second. In characterizing her solution to the issue of *de re* necessities, Thomasson relies on such a bifurcation between "semantic" knowledge and "empirical" knowledge, claiming "a modal normativist needn't and shouldn't be wedded to the idea that all necessities are knowable based solely on semantic competence," (111). On the account offered here, where modal normativism is applied not just to *metaphysical* necessities but also *nomological* necessities, it is maintained that all necessities *are* knowable through semantic competence. It's just that the notion semantic competence is expanded to include competence in scientific vocabularies, which have been structured through the process of conceptualization and reconceptualization through beholdenness to independent objects.

The second key point of contrast between this account and Thomasson's is that this account enables us to maintain a robust sense in which modal vocabulary, when it's deployed in the articulation of a scientific theory, really is describing the structure of independent reality. The crucial epistemological upshot of the previous five chapters, and emphasized by Thomasson as "the most important advantage of the normativist view" (147), is not compromised by this realist account of the relationship between language and the world that obtains in the case of specialized scientific practices. Even in the case of scientific practices, our grasp of properties is still, in the first instance, grasp of the

rules governing the use of linguistic expressions. Practices that are structured through the meta-practice of beholdenness to objects, such that the SG and ET relations depicted above obtain, *do* enable us to grasp the structure of independent reality. However, our grasp of this structure is not *direct*, as one might (mythically) imagine our grasp of the structure of sensible qualities such as redness to be, but *mediated* through one's grasp of the norms governing the expressions used in specialized linguistic practices. Articulating the rules governing the use of those expressions in a worldly mode, with modalized conditionals of the sort explicated in the previous chapter, one actually articulates aspects of the structure of reality. So, our grasp of the structure of independent reality is indeed obtainable—this account does not amount to any sort of skepticism about independent reality. However, grasp of the structure of independent reality is obtainable only through grasp of the structure of a linguistic practice which has been shaped through this process of conceptual revision.

6.2 Three Grades of Theoretical Status

On the account of properties we have provided, grasp of a property is always, in the first instance, grasp of the rules governing the use of linguistic expressions, and I claimed above that, crucially, this point applies even in the case of scientific practices. This is particularly clear when we consider physical theories that are formulated in the language of mathematics. The grasp of the properties of Newtonian mechanics, for instance those of *mass*, *force*, *velocity*, and so on comes by way of mastering the norms governing the use of the theoretical expressions "m," "F," "v," and so on, that figure in equations like $F = ma$, $a = \frac{\Delta v}{\Delta t}$, and so on. Our grasp of what these theoretical properties *are* is, in the first instance, is a reified grasp of the norms governing the theoretical terms in this mathematically regimented scientific vocabulary.[5] Now, of course, it is not *just* our grip of the abstract mathematical structure expressed by Netwon's equations that figures in

[5]The conception of mathematics that complements this account of meaning and the world, of course, is a structuralist one. The mathematical universe is a reification of the norms governing the use actual and possible bits of mathematical vocabulary, where this vocabulary is underlain by a structured set of actual and possible processes of construction, manipulation, and so on. This determines a vast space of possible mathematical structures, and some of these structures are actually instantiated in independent reality.

our grasp of these theoretical properties. Our initial grasp of these theoretical terms is also clearly tied to our grasp of the ordinary expressions such as "heavy," "massive," "push," "pull," "speed," "direction," and so on, though definitions of theoretical terms in terms of ordinary expressions such as "Something's velocity is its speed in a certain direction." Such ordinary explications of theoretical terms underlie our initial grip of the conceptual significance of the abstract mathematical representations of the theoretical properties, such as the vector representation of the velocity of an object. The point here is that our grasp of *both* sorts of concepts that figure in our grasp of theoretical terms, the mathematical concepts and the ordinary concepts, are to be understood in terms of our grasp of rules governing the use of expressions, mathematical or ordinary, and so our grasp of the theoretical properties of a physical theory is also to be understood, at least in the first instance, in terms of our grasp of the rules governing the use of linguistic expressions.

This essential connection between our grasp of the rules governing the use of ordinary expressions and our grip on theoretical properties is not lost even when we consider unobservable properties. Whereas velocity, a theoretical property that figures in Netwon's equations, can be observed, the electric and magentic fields and their properties and relations, which figure in Maxwell's equations, cannot be, at least, not directly. Even though the electric and magnetic fields are themselves unobservable, it can be very helpful to conceive these theoretically postulated fields by analogy to an observable fluid such as water in order to comprehend the significance of the theoretical expressions that figure into Maxwell's equations which articualte the constitutive standards of these fields. Consider the equation $\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$ which describes the divergence of a given region of the electric field as a function of the charge density of that region of the electric field. Mathematically, the electric field is understood as a vector field, a three-dimensional space that has a magnitude and direction at every point, and divergence is a mathematical concept that has its home in vector calculus. However, we can imagine the application of the concept of divergence at play here by conceiving of the electric field by analogy to a liquid like water, springing out from a source at some points and draining into a sink at others, where with the former providing the model for *positively* charged regions of the electric field and the

latter provides the model for *negatively* charged regions of the electric field. Of course, we know that the electric field is not really a *liquid*, but it functions sufficiently like a liquid in Maxwell's theory of electromagnetism that the application of the mathematical concepts of divergence and curl to the electric field it can be explicated by analogy to their application to a liquid like water: where we articulate water as a vector field, with the vectors at each point signifying the speed and direction of flow of the water, we have positive divergence where we have sources from which the water flows and negative divergence where we have sinks into which the water flows. In this way, the analogy to an observable phenomena such as the flow of water enables us to comprehend the application of the mathematical structure articulated by Maxwell's equations to the unobservable electric field. Once again, the point here is that, even in the case of unobservable theoretical properties, our grip on what they are is a product of our grasp of the rules governing the use of the mathematical expressions, on the one hand, and our grasp on the rules governing the use ordinary expressions, on the other. In this case, however, our grasp of the rules governing ordinary expressions confers substance on our grasp of mathematically-articulated theoretical terms only by analogy.

Now, it is certainly possible that we may get to a point in physical theorizing at which every analogy we might try to draw to provide substance to our mathematical formalism breaks down so fundamentally that we simply have to "shut up and calculate" (Mermin 1989). However, there is reason for optimism about our capacity to genuinely grasp the contents of even our most esoteric physical theories insofar as we take seriously Sellars's (1968) claim that "the use of analogy in theoretical science [. . .] generates new determinate concepts," (49). Insofar as this is so, one thing that we can do in making sense of physical theories is make analogy to other physical theories on which we have a grip, even if our initial grip on those other theories is *itself* only by analogy. For instance, in explicating quantum field theory to physics students who have not yet been initiated into it but who have been initiated into the theory of electromagneticism in prior courses, instructors feel free to bypass the analogy with observable fluids like water and to instead directly draw an analogy to such things as the electric and magnetic fields on which the students are already presumed to have a theoretical grip, as these theoretical objects provides

better model for thinking about the quantum fields than ordinary objects like water. This provides, in effect, a ladder of theoretical abstraction, and so the pool from which we can draw in providing analogies to provide substantive content to the mathematically-articulated theoretical concepts of our scientific theories actually grows in the course of theoretical inquiry. This enables the world articulated by our scientific theories to go quite far—indeed, *arbitrarily* far—from the world of common sense. So, though science is, as Sellars (1956) says, a "sophisticated extension" of common sense, it is ultimately utterly untethered to our common sense conception of things.

It is worth taking a moment at this point to explicitly note how the framework of discursive role semantics that has been proposed nicely complements this conception of theoretical freedom in scientific practices. First, it should be clear, at this point, that, though I've articulated constitutive standards in the section above as articulating *objects*, this basic conception of the engine of scientific inquiry in terms of beholdenness to objects is not necessarily tied to an ontology of discrete objects instantiating properties and standing in relations. The assumption that such an ontology is fundamental has been notably criticized, on empirical grounds, by Ladyman and Ross (2007) and French and Ladyman (2003), who argue that a proper understanding of quantum field theory, for instance, requires an interpretation of it according to which "the field is the structure, the whole structure and nothing but the structure. [. . .] [W]e can't describe the nature of the field without recourse to the mathematical structure of field theory," (48). Of course, this is not the place to get into a debate over the correct interpretation of quantum mechanics (not that I would have anything to say on the matter even if it was), but it's important to note that the conception of the objective world on offer here—as the alethic modal structure that corresponds to the normative structure of a scientific vocabulary shaped by beholdenness to the independent phenomena—is perfectly compatible with the ontology of the world being radically different than the familiar ontology of individual objects, properties that they instantiate, and relations that they stand in. Different categorial ontologies will simply correspond to different structures of discursive norms. Radically different structures may be necessary in the course of scientific inquiry, and there is very little limiting the way that our discursive practices must be structured. To give just one

example, note that, though we gave classical sequent rules in Chapter Four, it should be clear that any number of sequent systems can be implemented in this manner, for instance, those that determine various non-distributive quantum logics (see, for instance, Restall and Paoli 2005), with the alethically-articulated algebraic structure yielded by such rules (for instance, an Ortholattice, in this case) being the conception of the structure of the world conferred by discursive practices with such a structure.

6.3 The Emergence of "Two Worlds" and the Integrationist Task

As we have seen, the vocabularies produced by scientific theorizing have a relative autonomy with respect to our orindary linguistic practices. Through potentially iterable analogical reasoning and unbound mathematical regimentation, the structure and content of scientific practices is ultimately not teathered to that of ordinary linguistic practices. Still, these vocabularies do not generally stay isolated in the scientific practices that confer them. Rather, they often end up shaping our ordinary linguistic practices. As a result, semantic competence in ordinary English involves grasp of scorekeeping principles that originally came about through scientific inquiry and which, at one point, only shaped the structure of specilized scientific practices. A full merging of the norms governing the use of specialized scientific vocabularies and those governing the use of ordinary vocabularies, however, is a practical impossibility. Specialized scientific practices, which are structured by a distinctive set of orienting norms, are bound to maintain a certain sort of independence relative to ordinary language. Accordingly, two distinct pictures of reality, two "worlds," are bound to emerge: the "world" that is a reification of our ordinary vocabularies, and the world articulated by our scientific vocabularies, which are constitutiviely structured by a beholdenness to independent reality.

Before turning to the relative independence of ordinary and scientific practices, let us look first at how scientific practices can shape our ordinary ones. Consider a principle briefly mentioned earlier, that commitment "x is a whale" commits one to "x is a mammal" and thereby precludes one from being entitled to "x is a fish." This is something that a child learns in the course of elementary education, as part of the formation of the

concepts of different kinds of animals, and a speaker who takes commitment to "x is a mammal" to commit one to "x is a fish" can surely be regarded as less than fully competent with respect to this vocabulary. To be clear, the failure of competence here is to be understood as a failure of competence with respect to ordinary English, not a specialized scientific vocabulary. In the mid eighteenth century, however, the classification of whales as mammals was a development within a specialized scientific practice, made on the basis of anatomical comparison which was eventually taken provide a more systematic measure of classification than habitat.[6] Though this classification was initially controversial within the scientific community, it was eventually widely accepted by the late eighteenth century. Its widespread acceptance in scientific communities, however, predated its integration into ordinary linguistic practices, and the push to integrate this scorekeeping principle into ordinary linguistic practices was met with substantial resistance, crystallized in an 1818 court case concerning the question of whether whale oil fell under a taxation statute concerning "fish oil." This case ultimately ruled in favor of the "common sense" that whales, at least in the sense discussed in the letter of the law, were indeed fish (Sampson 1819). A similar court case occurred in 1886, once again due to taxation laws, concerning the question of whether tomatoes, considered fruits from a botanical perspective, were in fact fruits rather than vegetables, also ultimately ruling in favor of common sense. Surely, there are many other similar cases, but citing just these two is sufficient for an important point to be made.

As these two cases illustrate, there is no uniform way in which these sorts of conflicts between scientific vocabularies and ordinary vocabularies are always resolved. In the case of the whale, the scorekeeping principles constitutive of the biological kinds are generally those we are bound by in ordinary practice when we use the terms "fish," "whale," and "mammal." Someone who calls a whale a "fish" is making a mistake, displaying a lack of competence with the English language. This is not so with the case of someone who calls a cucumber a "vegetable." In the case of the cucumber, we now distinguish between two senses of the term "fruit," the *botanical* sense, on the one hand, and the *culinary and*

[6] In the ninth edition of *Systema Naturae*, Linnaeus (1756) notably first claimed that whales were to be classified as mammal, though it was met with resistance in among taxonomists at the time.

nutritional sense, on the other, maintaining that, while cucumbers are fruits in the first sense of "fruit," in the second sense of "fruit," they're not fruits but vegetables.[7] It is the scorekeeping principles constitutive of this second sense of "fruit"—according to which tomatoes, cucumbers, and peas, and are not fruits but vegeteables—that we generally bind ourselves by in ordinary practice when we use the term "fruit." So, in the first case, we have a merging the ordinary and the biological sense of the kind term, and, in the second case, we have a divergence of these two senses. I take it that there is no systematic way to determine, for a given term that has a use in both ordinary and scientific practices, whether these senses will end up merging or diverging.[8] Scientific vocabulary emerges out of natural language, and so terms like "whale," "fruit," "water," "gold," and "star," which have their initial home in ordinary practices, are used in scientific ones, but the use of these terms takes on a relative autonomy in the context of scientific practices, as those practices are subject to different pressures than ordinary ones, and so we have a potential divergence between the use of the term in the context of the scientific practice and the use in the context of the ordinary one. Sometimes, the senses of the terms used in respective practices end up reconverging, and sometimes they don't.

On the account of properties offered here, it is perfectly acceptable to say that there are properties grasped by speakers of a natural language that will never be identified with those articulated by a natural scientific theory. Indeed, presumably the vast majority of properties grasped by speakers of a natural language are such properties. Consider again the ordinary property of being a fruit, which is distinct from the botanical property of being a fruit. This is a property grasped by speakers of a natural language such as English, who grasp that apples and mangoes are fruits but cucumbers and peas are not fruits but vegetables, but it is not a property that exists as an aspect of independent reality. This property *exists*, as a reification of the discursive role of "fruit," but it exists as a *mere*

[7]This distinction underlied the ruling. "Botanically speaking tomatoes are the fruit of the vine, just as are cucumbers, squashes, beans and peas. But in the common language of the people... all these vegetables... are usually served at dinner in, with, or after the soup, fish, or meat, which constitute the principal part of the repast, and not, like fruits, generally as dessert."

[8]This can be seen as part of the moral of classic critiques of Putnam's scientific externalism such as those provided by Dupre (1981) and Laporte (1996). As Hacking (2007) puts this point, "Because of the sheer contingency we have no idea what we would say, let alone should say, if a Twin-Earth were ever to be discovered," (275).

reification: there is no bit of reality, independent of linguistic practices, that instantiates the bit of alethic modal structure constitutive of this property. That's just another way of saying that this property doesn't "carve nature at its joints," to use a common phrase. We might say that this property is uninstantiated, but, of course, we must acknowledge that there is a *sense* in which the property of being a fruit is instantiated. After all, it's instantiated by some things, such as apples and mangoes, and not others, such as cucumbers, peas, steaks, and couches. Saying such things, however, is just a way of expressing the norms governing the use of "fruit." So, though we can talk of the things that instantiate this property, expressing the norms of which it is a reification, we can nevertheless maintain that it's not *really* instantiated. That is to say, it's not a codification of a bit of reality, independent of the linguistic practice that confers it. Very many, indeed perhaps most, of the properties conferred by ordinary linguistic practices are like this, presumably such things as tables and chairs, stop signs and contracts, novellas and screenplays, and so on. This should not be too surprising of a statement; carving nature at the joints is simply not a principle aim of ordinary language. Its primary function is to enable us to get by in the world, given our interests, rather than describe the world, independent of our interests.

To further fill out the picture that's emerging here, let us consider one more example of a property whose story is significantly less straightforward than the ones we have just considered, one which we've taken as our main example here. Consider again the property of being red, which can be picked out by its location in the three-dimensional structure of value, hue, and saturation that was articulated in scorekeeping terms in Section 5.9. When we investigate the nature of red things, like stop signs and ripe tomatoes, and try to find some property that they have which instantiates this structure, we come up short. The only plausible candidate for a property that all of these things, as they are in themselves, actually share is a reflectance property: they all reflect light at a wavelength of around 700nm. However, if we try to identify colors with such properties, we see that they don't instantiate the structure that color properties essentially do (Pautz 2006). For instance, that it's essential to the property of being red that it is closer in hue to purple than it is to green, but the wavelengths of light reflected purple things are around 400nm and those

reflected by green things are around 550nm, so, by this identification, green should be closer to red than purple. The conclusion it seems that we are compelled to draw, on theoretical grounds, is that color properties are uninstantiated. However, unlike the case of the property of being a fruit just considered, it seems that the property of being red cannot be *merely* a reification of the discursive norms governing the use of "red." If it was a mere reification of discursive norms, then, plausibly, we'd be able to simply change them. But that doesn't seem to be the case. It's not simply an arbitrary decision that commitment to "x is red" precludes one from being entitled to "x is green," or that commitment to "x is red" and "y is pink" commits one to "x is darker than y." It seems that there is a reality underlying these norms, a reality revealed to us in color experience.

Now, here is the interesting fact about color: the structure of the three-dimensional space in which color properties can be located does correspond to *something* in reality, uncovered by scientific theorizing, but the bit of reality that instantiates this structure is a very different bit of reality than the bit of reality we pre-theoretically take to instantiate this structure. In particular, there are no properties instantiated by the external objects that we take to be colored that instantiate the structure of color properties; rather, it is internal states of ourselves that instantiate this structure. When light of different wavelengths hits our eyes, the pattern of activation of the three distinctive cones in our retinas can be understood in terms of three types of opposition, one between blue and green light, one between yellow and blue, and one between black and white. These three types of opposition correspond to three dimensions of a space of possible activation frequencies in the visual cortex which can be depicted as follows (Churchland 1995, 25):

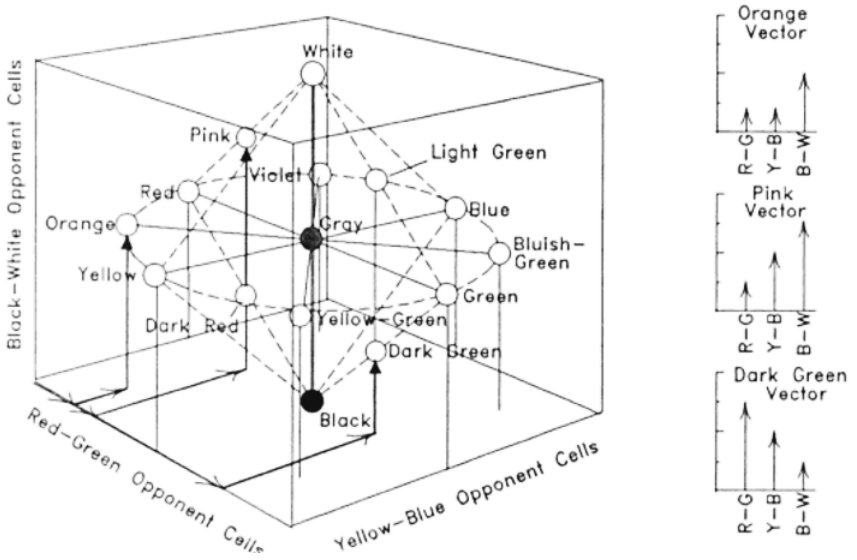


Figure 6.5: Sensory State Space w.r.t. Color

The structure of this space of possible states of activation in the visual cortex corresponds directly to the structure of the color space that we grasp first-personally in terms of the structured set of scorekeeping principles determining the discursive significance of expressions like "red," "green," "purple," "darker than," "closer in color to," and so on. The actual bit of reality underlying our having these scorekeeping principles is in fact not instantiated by external material objects such as ripe tomatoes and stop signs, but, rather, by internal states of ourselves. These internal states of ourselves, which we might speak of with the use of phrases such as "the state of sensing redly" or "the state of sensing greenly," are, of course, not themselves red or green, but they stand in relations of material entailment and incompatibility to one another that are analogous to the relations of material entailment and incompatibility that the color properties that we grasp stand to one another. It is through sensitivity to these states of ourselves, which we reliably and differentially enter into in virtue of looking at different kinds of objects that reliably reflect light at certain frequencies, that we are capable of learning color vocabulary and grasping the structure of color properties that exists as a reification of the norms governing the use of that vocabulary.[9] Discursive role semantics enables us to spell out the details of the complex relation that obtains between language and the world in this case.

[9]It might not come as a surprise that this is the story that Sellars (1968) suggests.

We are taught the colors by reference to observable material objects that are communally taken to instantiate color properties. That is, the linguistic norms that are actively enforced by the teachers of the language make reference to the intersubjectively perceptible properties of material objects, the redness of tomatoes and stop signs, for instance, that are recognized by the teachers who grasp the norms that govern the correct use of color vocabulary. So, when we become reflectively conscious of the norms that we've internalized through being brought into a linguistic practice, the things to which we take ourselves to be responsive, in using color expressions like "red," are visible properties of of objects like tomatoes and stop signs. Indeed, it's essential to the concepts that we acquire by being brought into a linguistic practice that involves the use of color expressions that these color expressions apply to intersubjectively accessible objects in the world. However, what we're *actually* responsive to, in using color expressions, are internal states of ourselves, our sensory states. The sort responsiveness to our own sensory states at play here is the sort of counterfactual relation described above as "epistemic tracking," where, if the sensory state were to be varied, the color judgment we are inclined to make would be systematically co-varied. The difference between the sort of epistemic tracking that occurs here and the sort of epistemic tracking discussed above in the context of our successful scientific vocabularies is that here there is a *mismatch* between the bits of the world that we're *epistemically tracking*, in using this vocabulary, and the bits of the world—or, rather, the "world"—that we take to *semantically govern* the use of the vocabulary. When we articulate the properties that exist as alethic modal reifications of the norms governing the use of color expressions, we articulate a space of *sensible qualities*, and that's what we take this bit of vocabulary to be representing, but what exists in reality that is actually responsible for the structure of our norms is the *sensory states* that we enter into in virtue of our trichromatic sensory system.

It turns out, then, that even the properties on which it seems we have the clearest and firmest grasp—color properties that seem to be manifestly instantiated by objects in our everyday experience—can be, and, in fact, *are*, uninstantiated. Acknowledging this fact does not bring with it any sort of commitment to drop talk of colors in our ordinary life. The sense of the reality of colors that we have as we go about everyday life is not one

that we could easily shake, nor is it one that we have any reason to shake, at least insofar as the ends of everyday life are concerned. There is no practical demand on us to revise our ordinary talk of the colors of things so that we no longer say, for instance, that we'd like orange throw pillows to go with the deep green sofa in the living room, and nothing prevents us from talking about the provocative effect of the girl in the red coat in *Schindler's List*. In this sense, there is nothing defective about our use of color vocabulary in everyday life. The world, as we intersubjectively experience it, is a world of colored objects, and so, as we go about our lives in the world that we share, the use of color vocabulary is perfectly apt. Still colors will not ultimately belong to the theoretical conception of the natural world that we achieve in scientific theorizing, except as reifications of discursive roles. What will belong our scientific conception, and what in part explains these discursive roles, are sensory states, reflectance properties, and various other things that are not themselves colored. The world, as it is in itself, contains no such things as colors.

From these considerations, a kind of "two worlds" picture emerges. On the one hand, there is the "world" of everyday objects, properties, and relations that are the reifications of the norms governing the use of the expressions of ordinary language. The properties appealed to in the context of worldly semantic theories belong to this world. Once again, these properties exist, the whole lot of them, but there is no guarantee that they exist as anything more than reifications of the norms governing the use of linguistic expressions. Now, in the previous chapters we raised a basic problem with worldly semantic theories: the "world," appealed to by these semantic theories as independent of linguistic norms, such that knowledge of it could underlie and explain of linguistic norms, is really to be understood as a reification of linguistic norms. One might have worried until this chapter that making this claim amounts to endorsing a problematic form of linguistic idealism. We can now see, however, that there is no commitment to linguistic idealism thrust upon us here, for this "world" that is a reification of the linguistic norms of our ordinary linguistic practices, can now be contrasted with the *world*, independent of our linguistic practices, that we uncover through scientific theorizing. As articulated above, scientific practice is normatively structured by a beholdenness to objects, which are what they are and do what they do independently of what we take them to be. In virtue of this fact, the

objects themselves can be seen as providing normative standards for a specialized linguistic practice of this sort. Thus, by conceiving of the norms that structure a scientific practice in reified terms, one is capable of conceiving of the world, as it is in itself, independently of our practices. These "two worlds," are essential to Sellars's philosophical system, and he most picturesquely discusses them in terms of two "images," what he (1962) called the "manifest image" and the "scientific image."[10] These two "images," we might say, continuing with the metaphor, are the projections of two sorts of vocabularies that we have, ordinary vocabulary and scientific vocabulary, each structured by a distinctive set of norms.

Now, Sellars (1962) famously characterized the aim of philosophy as "to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term," (1). On this characterization, it's clear that the philosopher's aspiration is to conceptually occupy *one* world that contains all the things, and to be able to conceptually navigate it and articulate what the things in it are and how they relate to one another. Thus, insofar as we have a "two worlds" picture, there is reason to want to integrate these worlds; to achieve, as Sellars puts it, a "stereoscopic vision" of the world, where the eye with the scientific image in view and the eye with the manifest in view jointly produce a single image. There are two integrationist tasks that must be completed in order to arrive at a unified picture of the world. The first is to unify the various domains *within* the scientific image, which is essentially incomplete insofar as it appears as several distinct images.[11] Various scientific disciplines are structured by the meta-practice of beholdenness to objects, but reifying the distinct vocabularies of the distinct disciplines yields a conception of the world with multiple distinct domains of objects without a particularly clear conception of how they are related. It is a commitment within scientific practice as a whole *that* these domains are systematically related, but one will not find an account ready at hand within scientific practice of just *how* they are. Providing such an account, Sellars thinks, is one of the principle tasks of philosophy, though he himself doesn't engage in this task as much as one might expect him to, given

[10]See Simonelli (2021, forthcoming) for a development of the idea of these two worlds in Sellars.
[11]See Hicks (2020) for a discussion of how this issue figures in Sellars's thinking.

this commitment. The other principle task of philosophy is articulating how the world articulated by a unified scientific practice will hang together with that articulated by our ordinary vocabularies. The first task, of unifying the various scientific disciplines, is a monumental task, one that cannot be undertaken but can only be gestured at here. That is what I will do in the next section. With the promissory note that such a task will need to be properly undertaken and achieved in due course, we'll then consider the second task of connecting the "world" of conferred by ordinary language with that articulated by science.

6.4 Towards a Unified Scientific Worldview

One of the orienting commitments of this dissertation (and, indeed, a basic point of agreement between the project undertaken here and the interdisciplinary work done in the semantic frameworks that I have criticized) is that language can itself be understood as a natural phenomenon, an object of natural scientific inquiry.[12] Given the account that I have developed, it follows that the"world" of ordinary language, with all the objects, properties, and relations in it, is itself also ultimately an object of scientific inquiry; the branch of scientific inquiry that articulates it is natural language semantics. I have argued that our basic understanding of language should be as a norm-governed social practice. Discursive role semantics involves the systematic articulation of the norms that determine the semantic significance of the various linguistic expressions belonging to a natural language, and the "world" of ordinary language is a reification of those norms. Unifying the "two worlds," then—integrating the world of ordinary language into the world of scientific theorizing—is nothing other than integrating our semantic theory into the rest of scientific theorizing. Moreover, insofar as natural language semantics is itself a science, this is just a part of the task of integrating our various scientific disciplines. Still, this is not a simple task. In explaining how language fits into the world, on a scientific conception of it, we must first have a scientific conception of the world that has

[12]That is not to say, of course, that this is the only way in which language can be understood. There is a reason why many sorts of inquiry into human language are understood as belonging to the "Humanities" rather than the "Natural Sciences." However, despite the division to which this dissertation has been submitted, the approach I am taking here aligns me more with the natural sciences.

enough conceptual resources in it to make sense of language as fitting into it, scientifically conceived. Specifically, what we need is an account of how the normative practices constitutive of a natural language such as English can be understood as emerging out of a world ultimately articulated in terms of such things as quantum fields, planets, carbon atoms, ribonucleaic acids, neurons, and so on. As I said, I will no more than gesture at this larger task of scientific unification, but some remarks about the commitments involved in taking there to be such a task that can be productively undertaken are necessary to get the overall picture I'm trying to sketch into view.

There is commitment here, shared with Sellars, to a certain sort of unity of science. This is best understood as a modal commitment concerning the *potential* for a unified picture: scientific theorizing, though certainly not *actually unified* at the moment, is in principle *unifiable*. In spelling out what this unification would come to, we can (at least nominally) tease apart two closely intertwined commitments here. The first commitment is to a sort of *material unity* among the sciences. That is, there is, fundamentally, one sort of "stuff" out of which all of objects of scientific inquiry are ultimately constituted. Insofar as we take physics to be the scientific discipline that articulates the fundamental structure of reality, this commitment to the material unity of the sciences amounts to a commitment to *physicalism*. This brings us to the second commitment, and that is to a sort *explanatory asymmetry* among the sciences, with one science articulating the fundamental level of explanation from which other levels of explanation can be understood as arising, where those levels of explanation, in turn, constituting levels of explanation out of which less fundamental levels can be understood as arising, and so on. Now, in criticizing Sellars's conception of a unified scientific image, Brandom (2015) writes "hardly any philosopher of science would subscribe to the explanatory hierarchy central to the unity-of-science idea," (85). As a sociological claim, this is simply false. Not only is some sort of explanatory hierarchy thesis held among the vast majority of serious philosophers of science, it is nearly ubiquitous among scientists themselves and can be regarded as an orienting commitment of science, structuring the interaction between the various disciplines in the actual practice of science.[13] Of course, hardly any philosopher of science these days would subscribe

[13]For instance, a recent discussion of the explanatory hierarchy (Rueger and McGivern 2010), begins

to the specific version of the explanatory hierarchy thesis influentially put forward by Oppenheim and Putnam (1958), which involved a rather strong form of reductionism that is rightly criticized in the classic articles that Bandom mentions in this regard such as those of Fodor (1974), (later) Putnam (1975), Dennett (1991).[14] However, throwing out the explanatory hierarchy thesis with reductionism would be to throw out the baby with the bathwater. To put this point more carefully, we can say that the joint commitment to the *material unity* and *explanatory asymmetry* of the sciences is compatible with a commitment to the *formal disunity* of the sciences. That is to say, there are forms of explanation present in, for instance, the biological sciences that cannot be understood in terms of the forms of explanation present in the physical and chemical sciences. So, the sort of *semantic reduction* of biological vocabulary to physical and chemical vocabulary, imagined in early the positivist versions of the hierarchy thesis, is not possible. Nevertheless, there is a sort of explanatory sufficiency relation that can be articulated between the vocabularies. Let me explain.

Drawing again on Brandom's (2008) own conceptual resources for articulating relationships between vocabularies, we can distinguish between different sorts of "sufficiency" relations between vocabularies. One sort of VV-sufficiency relation is *semantic reduction*. This is the sort of relation that is presupposed when one purports to give a definition of one expression in terms of others, saying, for instance, "A bachelor is an unmarried man" or "A prime number is a natural number greater than one that is not a product of two lesser natural numbers." Brandom conceives of the traditional project of philosophical analysis in terms of its attempts to carry out semantic reductions of this sort, showing, for instance, that the vocabulary of arithmetic is semantically reducible to the vocabulary of first-order quantificational logic with identity or that the vocabulary of ordinary objects is

with the sentence, "Talk of levels or layers of reality is ubiquitous in science and in philosophy. It is widely assumed, for instance, that physical, chemical, biological, and mental phenomena can be ordered in a hierarchy of levels, and that what happens at the so-called micro level determines the goings on at the macro," (379).

[14]Indeed, Dennett's "Real Patterns." has been widely appealed to in order to *explicate* the explanatory hierarchy thesis, rather than *refute* it. Dennett himself is clearly a proponent of some form the explanatory hierarchy thesis. As a matter of Sellars exegesis, Brandom doesn't note that Sellars himself was a strong proponent of emergence, and clearly not committed to the sort of reductionism targeted by these classic articles. Indeed, as Gabanni (2019) documents, one of one of Sellars's notable disputes with Carnap was over precisely this point.

semantically reducible to the vocabulary of sensations. In an attempt to reanimate the traditional project of analysis in response to seemingly fatal pragmatist criticisms, Brandom introduces a class of VV-sufficiency relations that are not *direct*, as semantic reductions are, but *mediated* by the specification of a set of *practices*: pragmatically mediated semantic relations. For instance, on Brandom's account, indexical vocabulary is not *semantically reducible* to non-indexical vocabulary, but non-indexical vocabulary can nevertheless be deployed to *specify* what one must *do* in order to be counted as correctly deploying indexical vocabulary.[15] What I want to suggest is that the relation between two scientific vocabularies that are not reducible to each other but where one stands in the asymmetric explanatory relation to the other appealed to in the unity of science conception is a similar sort of mediated VV-sufficiency. The mediation here, however, is not the mediation of linguistic practices, but of worldly states and processes through history. Specifically, this sort of sufficiency relation obtains when a base vocabulary V_1 is sufficient to *specify* a state of the world W_1, along with a set of rules according to which the states of the world progress, such that, with this state processing in accord with those rules through the course of history, another, differently structured, state of the world W_2 comes about that suffices for the *applicability* of a target vocabulary V_2. The resultant vocabulary-vocabulary sufficiency relation is the composition of these three relations. Using our worldly extension of Brandom's meaning-use diagrams once again, we might depict this set of relations as follows:

[15] I should be clear that I do not intend to endorse this claim about indexical vocabulary; I am merely using it as an example to illustrate Brandom's notion of a pragmatically mediated semantic relation.

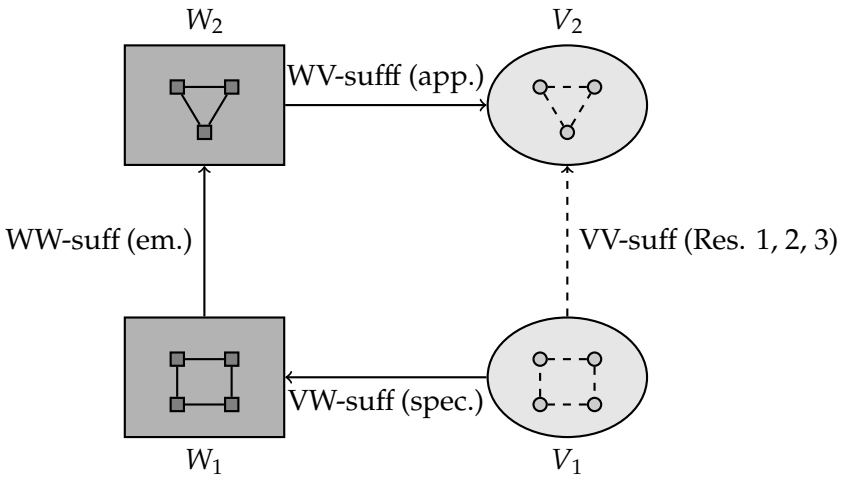


Figure 6.6: Emergence Mediated VV-suff Relations

Where this set of relations obtains, I'll say that we have a VV-sufficency relation mediated by emergence.[16]

It is the obtaining of this sort of non-reductive VV-sufficency relation, I believe, that underwrites the applicability of the *transformative* conception of life, animality, and discursive rationality that Mathew Boyle (2012) puts forth in his paper "Essentially Rational Animals." This transformative conception applies to the major transitions that took place through course of natural history, from a world of merely natural things, to a world of living things whose constitutive standards include their being oriented towards the end of preserving themselves as the living things that they are, to a world of perceptive and active living animals that navigate their environment in a way that is rationally intelligible to us, and finally, us who not only operate in accord with instrumental norms of rationality in our own activity but hold one another and ourselves to shared discursive norms that we can explicitly articulate.[17] With each of these transitions, a distinctive vocabulary with a distinctive formal—or we might even say *logical*—structure comes to be applicable, from merely alethic modal vocabulary, to teleological vocabulary, to the vocabulary of

[16]I take it that the WW-sufficiency relation expressed here is a kind of *grounding* relation. As I will understand it here, grounding is what Karen Bennett (2011) dubs a "super-internal" relation, where the intrinstic nature of just one of the relata suffices both for the relation that obtains between it and the other relata and for the other relata itself.

[17]For a more substantive development of this idea of change in form from living, to instrumentally rational, to discursive beings, see especially the work of Mark Okrent (2001).

instrumental rationality, and, finally, the sort of social normative vocabulary that I have deployed throughout the positive portion of this dissertation: the vocabulary of *commitments* and *entitlements*. Each of these vocabularies, I claim, stands in this sort of mediated VV-sufficiency relation to the next; none is semantically reducible to the others, but the applicability of each of the posterior vocabularies can be explained through the deployment of the prior vocabularies. Actually articulating these mediated VV-sufficiency relations can be understood as *the* project of philosophical and scientific naturalism. Once again, anything more a gesture at this project is beyond the scope of the current work, but I hope I have at least made clear that there is such a project to be undertaken.

6.5 Explaining Human Language

The major transition in the course of natural history that is of the most immediate concern for the purposes of the present project is the one from *instrumentally* rational animals, such as chimpanzees, to *discursively* rational animals, such as ourselves. Brandom (1994) describes us as belonging to this latter category by saying, "We are the ones on whom reasons are binding, who are subject to the peculiar force of the better reason," (5). The notion of "reason" of which Brandom speaks here is essentially *intersubjective*, something that can be *given* to others or *called for* by others. In engaging in the discursive practices that we do, we take ourselves to be bound by the reasons given to us by others, and we take ourselves to be bound by the demand to give reasons when others call upon us to do so. This is not something that our closest hominid relatives, the chimpanzees and bonobos, do, as evidenced by a mountain of research undertaken most substantially by Michael Tomassello and his colleagues (2003, 2008, 2014). The core idea unifying the data that Tomassello and his colleagues have uncovered is that only in the context of cooperative activity, where we're mutually directed towards some shared end that we both recognize as our shared are, can we get the basic sort of normative infrastructure that can eventually developed into the complex conventionalized normative infrastructures that are the languages we have today. [18] Now, the *phylogenetic* question, of how the sort

[18] This connections of between this work by Tomassello and his colleagues, articulating the emergence of the normative social practices that fundamentally distinguish us from our closest homind ancestors, and the

of normative social practices through in which discursive agents are constituted come to be through the course of natural history, is intimately intertwined with the *ontogenetic* question of how a particular discursive-agent-to-be comes to be what it is through being brought into a discursive practice. The reason these two questions are intertwined is that discursive practices are self-consciously shaped by the practitioners that collectively constitute them. That, is once one is constituted as a discursive being, one can deploy one's discursive capacities to actively shape the discursive practice into which future discursive beings will come to be constituted, and so on. Thus, while it's true, as Wittgenstein tells us, that, when one learns a language, one does so blindly, the language that one thereby learns is anything but blindly constructed.

The logic of the ontogeny of linguistic understanding is articulated by Sellars (1969) in terms of a potential language speaker's, in the first instance, having their performances being brought into conformity to rules of *criticism*—what he calls "ought to be" rules—and eventually holding themselves to corresponding rules of *action*—"ought to do" rules.[19] For instance, bringing someone into an English-speaking practice involves brining their behavior into conformity with rule that one ought to be such that one responds to red things by saying "red" in appropriate circumstances, the rule that one ought to be such that one does not say both "red" and "green" in application to some thing, and so on. These "ought to be" rules, are underwritten by corresponding "ought to do" rules consciously followed on the part of the teachers, for instance, that one ought to encourage responses of "red" to red things in appropriate circumstances, and discourage applications of "green" to something to which "red" has been applied, and so on. The teachers of course, grasp the "ought to do" rules that one ought to say "red" in response to red things, if appropriately prompted, one ought not say "red" and "green" in application to the same thing, and so on. One comes to be a discursive being through being brought into a linguistic practice,

Sellarsian and Brandomian conception of language and discursive cognition I have advanced here has been recently been substantially developed by Preston Stovall (2022) as part of a larger project on naturalizing inferentialism. It's worth noting that, though the overall orientation of the work is quite close to that of the present work, the specific semantic framework developed by Stovall is one

[19] In understanding Sellars's distinction, one shouldn't get too hung up on the grammar of the locutions used to the express the respective kinds of rules. The grammatical distinction suggests the conceptual distinction, but one can, for instance, the language of "ought to ϕ" to express an "ought to be" rule, if the context is right.

being held to the norms, and, eventually, holding oneself to them. A conscious grip on the norms is a product of what Tomasello (2014) calls "normative self-monitoring" (118-120), by which one comes to regulate one's own performances according to the norms of the practice into which one is being brought, to which one is being held. Consciousness, in the distinctive discursive sense that human beings have it, is a product of this sort of normative self-monitoring.

Of course, the details of this story will need to be spelled out, but, when they are, this story will amount to a *transformative* conception of human consciousness and thinking (Boyle 2016). What distinguishes human thinking from that of non-human animal is that human thinking is *essentially discursively articulable*. The sort of thinking that is characteristic of human beings is such that, when one thinks something, one can articulate just what it is that one is thinking. With language, not only can express one's thoughts, but one can clarify one's own expressions of one's thoughts, saying just what one means, articulating precisely what commitments one takes oneself to be undertaking in saying what one does.[20] This capacity, which comes only with language, transforms what it is to be a thinking being. So, though non-linguistic animals may be said to be able to "think," in some sense of the term, the very form of the capacity that we ascribe to them in saying of them that they can "think" is distinct from the capacity that distinguishes us as discursive beings, beings who are capable of discursively articulable thought. Discursively articulable thought, precisely in virtue of its potential discursive articulation, is *determinate* in a way that thought of non-discursive animals is not. The very possibility of having a conscious grip on a thought that one has, of grasping just what it is that one thinks, depends on the capacity for linguistic articulation. It is in this sense that Sellars's "psychological nomianlism" should be understood. As Sellars says, "all awareness [. . .] is a linguistic affair," but, in saying this, we must clarify that, by "awareness" we mean the sort of awareness possessed by discursive beings, wherein one is capable of knowing what it is of which one is aware, getting a grip on the content of the awareness. So construed, psychological nominalism is ultimately truistic—indeed, tautologous in a way—boiling

[20]For a discussion of this sort of clarification of expression through linguistic means, see Finkelstein (2019).

down to the claim that the distinctive sort of awareness that is had by linguistic beings is always a linguistic affair! This, perhaps, should not be too surprising given that, if there is anything that is a general response to the Myth of the Given, it is psychological nominalism, and, as Conant (M.S.) tells us, if we try to say what the Myth in general is "we are bound to end up saying something that essentially has the form of the negation of a tautology."

This brings us back to the point at which this dissertation started: the claim that worldly semantic theories fall prey to the Myth of the Given. Throughout this dissertation, I have argued that the worldly knowledge to which such semantic theories appeal as underlying the capacity to speak a language can really be understood only as a product of that capacity. Even if these arguments seem good, however, one might think that they cannot actually go through, since this claim may seem to imply that animals who lack the capacity to speak a language don't have knowledge of the world, and, given that they go about the world navigating it in a way that clearly manifests awareness of the various things in it, such a claim seems palpably implausible. It should now be clear, however, that this claim is perfectly compatible with saying that there is a *kind* of knowledge of the world that is had by non-linguistic animals; it's just that that kind of knowledge is fundamentally distinct from the determinate knowledge that is attributed to us by a worldly semantic theory, which can in principle be discursively articulated in the way that I have demonstrated for the toy language we've considered. *This* sort of knowledge, which we might equally characterize as distinctively *conceptual* knowledge, can only be understood as a product of the capacity to speak a language. We have now given an account of just what it is in which that worldly knowledge consists. The question of how, exactly, to characterize the kind of knowledge possessed by non-linguistic animals and the further question of how, exactly, to articulate the transformative relation between the that kind of knowledge and the kind of knowledge we possess is an important one and one that falls beyond the scope of the current project. I hope I have done enough, however, to sketch the general shape of the view language and it's place in the natural world of which the of account of meaning and discursive knowledge that I've articulated here can be seen as a part.

6.6 Conclusion

I have argued for a fundamentally different approach to linguistic meaning than the approach that dominates contemporary semantic theorizing. Rather than presuming that speakers have knowledge of the world and attempting to explain their knowledge of linguistic meaning as asymmetrically depending on this worldly knowledge, I have proposed a semantic theory that enables us to think of things as going in the opposite direction. On the account I have proposed, it is only through being brought into a linguistic practice, being held the norms of those practice by others and eventually holding oneself to them, that one comes to have a grip on the "worldly" entities appealed to in worldly semantic theories. In this chapter, I have argued that, though our grip on the worldly entities is always, in the first instance, a reified grip on discursive roles, the entities on which we have a grip need not always be construed as *mere* reifications of discursive roles. In particular, in discursive practices that are structured by a beholdenness to objects, which are constitutely restructuring themselves in response to what the objects do, are such that, by having a reified grip on their norms, one thereby has a grip on the structure of objective reality. Knowledge of the world, as it is in itself, is not simply given to speakers of a natural language. Rather, it must be achieved through the active shaping of our language. Moreover, genuine self-knowledge, knowledge of who we really are, can be achieved only by finding ourselves in the world, as it really is, integrating our conception of ourselves into a scientific worldview. I hope this dissertation constitutes a step towards that ultimate end.

A

Supra-Classicality of the Sequent System

In Chapter Four, I proposed a bilateral sequent calculus, interpreted as a set of rules for expanding a set of scorekeeping principles relating commitments and preclusions of entitlement to *atomic* sentences to a set of scorekeeping principles relating commitments and preclusions of entitlements to *logically complex* sentences. For reasons that will be clear shortly, I call that system NM-B. Following the ROLE approach (Section 5.3), we might technically think of it as a calculus that extends a bilateral *base* consequence relation, $\vdash_0$, relating sequences of signed atomic formulas (on the left) to single signed atomic formulas (on the right), to a bilateral *extended* consequence relation, $\vdash$, relating sequences of signed formulas of arbitrary logical complexity. Where A, B, and C are any signed formula, Γ and Δ are sets of signed formula, and staring a signed formula yields the oppositely signed formula, we may specify this logic (with the material conditional added) as follows:

NM-B

Axioms Schemas:

$$\frac{}{\Gamma \vdash A}\ \text{MB}\ \text{if}\ \Gamma \vdash_0 A \qquad\qquad \frac{}{\Gamma, A \vdash A}\ \text{CO}$$

Where Γ and $\{A\}$ contain only signed atomics.

Structural Rules:

$$\frac{\Gamma, A, B, \Delta \vdash C}{\Gamma, B, A, \Delta \vdash C}\ \text{P} \qquad\qquad \frac{\Gamma, A \vdash B}{\Gamma, B^* \vdash A^*}\ \text{RV}$$

Operational Rules:

$$\frac{\Gamma \vdash \ominus\langle\varphi\rangle}{\Gamma \vdash \oplus\langle\neg\varphi\rangle}\ \oplus_{\neg} \qquad \frac{\Gamma \vdash \oplus\langle\varphi\rangle}{\Gamma \vdash \ominus\langle\neg\varphi\rangle}\ \ominus_{\neg}$$

$$\frac{\Gamma \vdash \oplus\langle\varphi\rangle \quad \Gamma \vdash \oplus\langle\psi\rangle}{\Gamma \vdash \oplus\langle\varphi \wedge \psi\rangle}\ \oplus_{\wedge} \qquad \frac{\Gamma, \oplus\langle\varphi\rangle \vdash \ominus\langle\psi\rangle}{\Gamma \vdash \ominus\langle\varphi \wedge \psi\rangle}\ \ominus_{\wedge}$$

$$\frac{\Gamma, \ominus\langle\varphi\rangle \vdash \oplus\langle\psi\rangle}{\Gamma \vdash \oplus\langle\varphi \vee \psi\rangle}\ \oplus_{\vee} \qquad \frac{\Gamma \vdash \ominus\langle\varphi\rangle \quad \Gamma \vdash \ominus\langle\psi\rangle}{\Gamma \vdash \ominus\langle\varphi \vee \psi\rangle}\ \ominus_{\vee}$$

$$\frac{\Gamma, \oplus\langle\varphi\rangle \vdash \oplus\langle\psi\rangle}{\Gamma \vdash \oplus\langle\varphi \supset \psi\rangle}\ \oplus_{\supset} \qquad \frac{\Gamma \vdash \oplus\langle\varphi\rangle \quad \Gamma \vdash \ominus\langle\psi\rangle}{\Gamma \vdash \ominus\langle\varphi \supset \psi\rangle}\ \ominus_{\supset}$$

The "B" in the name "NM-B" is, unsurprisingly, for "Bilateral." The "NM," however, is for *non-montonic*. What is notable about this sequent calculus is that it does not require the usual structural rule of Monotonicity to operate. While this structural rule posed no issues for our simple to language, it does pose issues insofar as we consider the set of scorekeeping principles constituitive of the meanings of actual ordinary expressions. For instance, in articulating the meaning of "bird," we will presumably want the principle $\oplus\langle\textbf{bird}\rangle \vdash \oplus\langle\textbf{flies}\rangle$, but not $\oplus\langle\textbf{bird}\rangle, \oplus\langle\textbf{penguin}\rangle \vdash \oplus\langle\textbf{flies}\rangle$. This calculus is capable of working to extend such a set of atomic scorekeeping principles. In other work (Simonelli M.S.b, M.S.d), I show how to define updates in such a way as to not impose this structural rule. Here, to supplement the claims made in Chapter Four, the only thing that needs to be shown is that this system suffices for classical logic, generating an extension of a classical consequence relation.

To show this, it is sufficient to show that this system is a bilateral twin of the system NM-MS (non-monotonic multi-succicent), proposed by Kaplan (2018), that adds non-logical axioms to Kentonen's (1944) classical sequent rules. Here it is:

NM-MS:

Axioms:

$$\frac{}{\Gamma \vdash \Delta}\ {}^{\text{MB}}\ \text{If } \Gamma \vdash_0 \Delta \qquad\qquad \frac{}{\Gamma, p \vdash p, \Delta}\ {}^{\text{CO}}$$

Where Γ and Δ contain only atomics.

Structural Rules:

$$\frac{\Gamma, \psi, \varphi, \Delta \vdash \Theta}{\Gamma, \varphi, \psi, \Delta \vdash \Theta}\ \mathrm{P}_L \qquad\qquad \frac{\Gamma \vdash \Delta, \psi, \varphi, \Theta}{\Gamma \vdash \Delta, \varphi, \psi, \Theta}\ \mathrm{P}_R$$

Operational Rules:

$$\frac{\Gamma \vdash \varphi, \Delta}{\Gamma, \neg\varphi \vdash \Delta}\ \mathrm{L}_\neg \qquad\qquad \frac{\Gamma, \varphi \vdash \Delta}{\Gamma \vdash \neg\varphi, \Delta}\ \mathrm{R}_\neg$$

$$\frac{\Gamma, \varphi, \psi \vdash \Delta}{\Gamma, \varphi \wedge \psi \vdash \Delta}\ \mathrm{L}_\wedge \qquad\qquad \frac{\Gamma \vdash \varphi, \Delta \quad \Gamma \vdash \psi, \Delta}{\Gamma \vdash \varphi \wedge \psi, \Delta}\ \mathrm{R}_\wedge$$

$$\frac{\Gamma, \varphi \vdash \Delta \quad \Gamma, \psi \vdash \Delta}{\Gamma, \varphi \vee \psi \vdash \Delta}\ \mathrm{L}_\vee \qquad\qquad \frac{\Gamma \vdash \varphi, \psi, \Delta}{\Gamma \vdash \varphi \vee \psi, \Delta}\ \mathrm{R}_\vee$$

$$\frac{\Gamma \vdash \varphi, \Delta \quad \Gamma, \psi \vdash \Delta}{\Gamma, \varphi \supset \psi \vdash \Delta}\ \mathrm{L}_\supset \qquad\qquad \frac{\Gamma, \varphi \vdash \psi, \Delta}{\Gamma \vdash \varphi \supset \psi, \Delta}\ \mathrm{R}_\supset$$

This logic is known to be supra-classical in that every classical entailment is contained within its consequence relation: this is just the fragment of $\vdash$ generated by proofs whose leaves only included instances of CO (Kaplan 2018, 8).

To show that NM-B is equivalent to NM-MS, and thus, supra-classical, we start with a translation schema relating any NM-MS sequent to an equivalence class of NM-B sequents:[1]

> **Translation Schema:** Let us write $\oplus\langle\Gamma\rangle$ to express $\oplus\langle\gamma_1\rangle, \oplus\langle\gamma_2\rangle \ldots$ for all $\gamma \in \Gamma$. Likewise for $\ominus\langle\Gamma\rangle$. The equivalence class of NM-B seqeunts (under Reversal and Exchange) for a NM-MS sequent of the form $\Gamma \vdash \Delta$ is the union of
>
> $$\{\oplus\langle\Gamma'\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\gamma\rangle \mid \Gamma' = \Gamma \setminus \{\gamma\} \text{ with } \gamma \in \Gamma\}$$
>
> and

[1] Many to Dan Kaplan for suggesting this way of formulating the translation schema.

$$\{\oplus\langle\Gamma\rangle, \ominus\langle\Delta'\rangle \vdash \oplus\langle\delta\rangle \mid \Delta' = \Delta \setminus \{\delta\} \text{ with } \delta \in \Delta\}$$

What we now need to show is that $\Gamma \vdash_{\text{NM-MS}} \Delta$ just in case $\Gamma^* \vdash_{\text{NM-B}} \Delta^*$, where the latter is shorthand for any of the equivalent NM-B translations of $\Gamma \vdash_{\text{NM-MS}} \Delta$. To do this, we do an induction on proof height to show that any sequent we get through an NM-MS proof is one whos translation we are able to get through a corresponding NM-B proof, and vice versa.

The base case (proof height = 1) is given directly by the translation procedure. Any NM-MS CO-instance of the form $\Gamma, p \vdash p, \Delta$ corresponds to the (equivalent) NM-B CO-instances $+\langle\Gamma\rangle, \ominus\langle\Delta\rangle, \oplus\langle p\rangle \vdash \oplus\langle p\rangle$ and $+\langle\Gamma\rangle, \ominus\langle\Delta\rangle, \ominus\langle p\rangle \vdash \ominus\langle p\rangle$ and vice versa. Any NM-MS material axiom of the form $\Gamma \vdash \Delta$ corresponds to the NM-B sequents given by the translation schema and vice versa.

For the inductive step, we suppose that the result holds for proof height n and show that it holds for proof of height $n + 1$. I will just show that the inductive step holds for the negation rules and the conjunction rules, as the disjunction and conditional rules are directly analogous to the conjunction rules. So, there are eight cases we need to consider to show that our inductive step holds.

Suppose the last step of a NM-MS proof is $\text{L}\neg$. Then the sequent at proof height n is of the form $\Gamma \vdash \varphi, \Delta$, and the sequent at proof height $n + 1$ is of the form $\Gamma, \neg\varphi \vdash \Delta$. By our inductive hypothesis, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi\rangle$, and via $\ominus_\neg$, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\neg\varphi\rangle$, which is a translation of $\Gamma, \neg\varphi \vdash \Delta$.

Suppose that the last step of a NM-B proof is $\ominus_\neg$. Then the sequent at proof height n is of the form $\oplus\langle\Gamma\rangle \ominus \langle\Delta\rangle \vdash \oplus\langle\varphi\rangle$, and the sequent at proof height $n + 1$ is of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\neg\varphi\rangle$. By our inductive hypothesis, we have $\Gamma \vdash \Delta, \varphi$, and by $\text{L}\neg$, we have $\Gamma, \neg\varphi \vdash \Delta$, of which $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\neg\varphi\rangle$ is a translation.

Suppose the last step of an NM-MS proof is $\text{R}\neg$. Then the sequent at proof height n is of the form $\Gamma, \varphi \vdash \Delta$, and the sequent at proof height $n + 1$ is of the form $\Gamma \vdash \neg\varphi, \Delta$. By our inductive hypothesis, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\varphi\rangle$, and via $\oplus_\neg$, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\neg\varphi\rangle$, which is a translation of $\Gamma \vdash \neg\varphi, \Delta$.

Suppose that the last step of a NM-B proof is $\oplus_\neg$. Then the sequent at proof height

n is of the form $\oplus\langle\Gamma\rangle \ominus \langle\Delta\rangle \vdash \ominus\langle\varphi\rangle$, and the sequent at proof height $n+1$ is of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\neg\varphi\rangle$. By our inductive hypothesis, we have $\Gamma, \varphi \vdash \Delta$, and by $R_\neg$, we have $\Gamma \vdash \neg\varphi, \Delta$, of which $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\neg\varphi\rangle$ is a translation.

Suppose the last step of an NM-MS proof is $L_\wedge$. Then the sequent at proof height n is of the form $\Gamma, \varphi, \psi \vdash \Delta$ and the sequent at proof height $n+1$ is of the form $\Gamma, \varphi \wedge \psi \vdash \Delta$. By our inductive hypothesis, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle, \oplus\langle\varphi\rangle \vdash \ominus\langle\psi\rangle$, and by $\ominus_\wedge$, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\varphi \wedge \psi\rangle$, which is a translation of $\Gamma, \varphi \wedge \psi \vdash \Delta$.

Suppose the last step of an NM-B proof is $\ominus_\wedge$. Then the sequent of proof height n is of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle, \oplus\langle\varphi\rangle \vdash \ominus\langle\psi\rangle$ and the sequent of proof height $n+1$ is of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\varphi \wedge \psi\rangle$. By our inductive hypothesis, we have $\Gamma, \varphi, \psi \vdash \Delta$, and by $L_\wedge$, we have $\Gamma, \varphi \wedge \psi \vdash \Delta$ of which $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \ominus\langle\varphi \wedge \psi\rangle$ is a translation.

Suppose the last step of an NM-MS proof is $R_\wedge$. Then the sequents at proof height n are of the form $\Gamma \vdash \varphi, \Delta$ and $\Gamma \vdash \psi, \Delta$, and the sequent at proof height $n+1$ is of the form $\Gamma \vdash \varphi \wedge \psi, \Delta$. By our inductive hypothesis we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi\rangle$ and $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\psi\rangle$, and by $\oplus_\wedge$, we have $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi \wedge \psi\rangle$ which is a translation of $\Gamma \vdash \varphi \wedge \psi, \Delta$.

Suppose the last step of an NM-B proof is $\oplus_\wedge$. Then the sequents at proof height n are of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\psi\rangle$ and $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi\rangle$, and the sequent at proof height $n+1$ is of the form $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi \wedge \psi\rangle$. By our inductive hypothesis, we have $\Gamma \vdash \varphi, \Delta$ and $\Gamma \vdash \psi, \Delta$, and, by $R_\wedge$, we have $\Gamma \vdash \varphi \wedge \psi, \Delta$ of which $\oplus\langle\Gamma\rangle, \ominus\langle\Delta\rangle \vdash \oplus\langle\varphi \wedge \psi\rangle$ is a translation. □

It follows that, like NM-MS, NM-B is supra-classical in that the entailments of classical logic are given by the fragment of the consequence relation generated by proofs whose leaves are instances of CO.

Bibliography

[1] Adams, Robert. 1974. "Theories of Actuality." *Nous* 8, no. 3 (September): 211-231.

[2] Armstrong, D. M. 1989. *Universals: An Opinionated Introduction*. Westview Press.

[3] Backstrom, Stina. 2016. "A Dilemma for Neo-Expressivism—And How to Resolve It." *Acta Analytica* 31, no. 2 (June): 191-205.

[4] Belnap, Nuel. 1990. "Declaratives Are Not Enough." *Philosophical Studies* 59, no. 1: 1-30.

[5] Bennett, Karen. 2011. "By Our Bootstraps." *Philosophical Perspectives* 25: 27-41.

[6] Berger, Jacob. 2015. "The Sensory Content of Perceptual Experience." *Pacific Philosophical Quarterly* 96: 446-468.

[7] Berger, Jacob. 2021. "Quality-Space Functionalism about Color." *Journal of Philosophy* 118, no. 3:138-164.

[8] Berto, Francesco. 2008. "*Αδνυατον* and material exclusion." *Australasian Journal of Philosophy* 86, no. 2:165-190.

[9] Berto, Francesco. 2015. "A Modality Called 'Negation'." *Mind* 124, no. 495: 761-793.

[10] Borschev, Vladimir and Barbara H. Partee. 1998. "Formal and lexical semantics and the genitive in negated existential sentences in Russian." In, Steven Franks and William Snyder, eds., *Formal Approaches to Slavic Linguistics 6*: The Connecticut Meeting 1997, Ann Arbor: Michigan Slavic Publications, 75-96.

[11] Boyle, Mathew. 2012. "Essentially Rational Animals." In *Rethinking Epistemology*, ed. G. Abel and J. Conant, 395-427. Berlin: de Gruyter.

[12] Boyle, Mathew. 2016. "Additive Theories of Rationality: A Critique." *European Journal of Philosophy* 24, no. 3: 527-555.

[13] Brandom, Robert. 1984. "Reference Explained Away." *The Journal of Philosophy* 81, no. 9: 469-492.

[14] Brandom, Robert. 1994. *Making It Explicit.* Cambridge, MA: Harvard University Press.

[15] Brandom, Robert. 1997. "Study Guide." In Wilfrid Sellars, *Empiricism and the Philosophy of Mind.* Cambridge, MA: Harvard University Press.

[16] Brandom, Robert. 2000. *Articulating Reasons.* Cambridge, MA: Harvard University Press.

[17] Brandom, Robert. 2008. *Between Saying and Doing.* Oxford: Oxford University Press.

[18] Brandom, Robert. 2010. "Response to Rebecca Kukla and Mark Lance." In *Reading Brandom: On Making It Explicit*, ed. by J. Wanderer and B. Weis. New York: Rutledge.

[19] Brandom, Robert. 2018. "From Logical Expressivism to Expressivists Logics: Sketch of a Program and Some Implementations." In *From Rules to Meanings: New Essays on Inferentialism*, ed. O. Beran, V. Kolman, and L. Koren, 151-164. New York: Routledge.

[20] Brandom, Robert. 2019a. *A Spirit of Trust.* Cambridge, MA: Harvard University Press.

[21] Brandom, Robert. 2019b. "Notes for Week 9: Alethic Modality II." Lecture Notes for *The Philosophy of Wilfrid Sellars* Phil. 2410. University of Pittsburgh, delivered October 23, 2019.

[22] Bridges, Jason. 2006. "Does Informational Semantics Commit Euthyphro's Fallacy." *Nous* 40, no. 3: 522-547.

[23] Burgess, Alexis and Brett Sherman. 2014. "Introduction: A Plea for the Metaphysics of Meaning." In *Metasemantics: New Essays on the Foundations of Meaning*, edited by A. Burgess and B. Sherman, 1-16. Oxford: Oxford University Press.

[24] Carnap, Rudolph. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.

[25] Carnap, Rudolph. 1952. "Meaning Postulates." *Philosophical Studies* 3, no. 5: 65-73.

[26] Cappelen, Herman and Ernie Lepore. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden: Blackwell.

[27] Chalmers, David. 2010. *The Character of Consciousness.* Oxford: Oxford University Press.

[28] Chierchia, Gennarro and Sally McConnell-Ginet. 1990. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.

[29] Churchland, Paul. 1995. *The Engine of Reason, the Seat of the Soul.* Cambridge, MA: MIT Press.

[30] Conant, James. 2020. "Replies." In *The Logical Alien: Conant and His Critics*, ed. Sophia Miguens, 321-1028. Cambridge, MA: Harvard University Press.

[31] Conant, James. M.S. "Sellars on the Dialectic of the Given." Presented at the University of Erlangen, Germany. June 13, 2015.

[32] De, Michael and Hitoshi Omori. 2018. "There is More to Negation than Modality." *Journal of Philosophical Logic* 47: 281-299.

[33] Dennett, Daniel. 1991. "Real Patterns." *Journal of Philosophy* 88, no. 1: 27-51.

[34] Dever, Josh. 2012. "Formal Semantics." In *The Continuum Companion to the Philosophy of Language*, edited by Manuel García-Carpintero and Max Kölbel. Continuum International.

[35] deVries, Willem and Timm Triplett. *Knowledge, Mind, and the Given: Reading Wilfrid Sellars's "Empiricism and the Philosophy of Mind."* Indianapolis: Hacket.

[36] Dowty, David, Robert Wall, and Stanley E. Peters. 1981. *Introduction to Montague Semantics.* Reidel.

[37] Dummett, Michael. 1991. *The Logical Basis of Metaphysics.* Cambridge, MA: Harvard University Press.

[38] Dunn, Michael. 1993. "Star and Perp: Two Treatments of Negation." *Philosophical Perspectives 7: Language and Logic*, ed. J. Tomberli, 331-357. Atascadero CA: Ridgeview.

[39] Dunn, Michael. 1999. "A Comparative Study of Various Model-Theoretic Treatments of Negation: A History of Formal Negation." In *What Is Negation?*, ed. D. M. Gabbay and H. Wansing, 23-51. Dordrecht: Springer.

[40] Dupré, John. "Natural Kinds and Biological Taxa." *The Philosophical Review* 90, no. 1: 66-90.

[41] Egan, Andy. 2004. "Second-Order Predication and the Metaphysics of Properties." *Australasian Journal of Philosophy* 82, no. 1 (March): 48-66.

[42] Fine, Kit. 2017. "Truthmaker Semantics." In *A Companion to the Philosophy of Language: Second Edition*, ed. B. Hale, C. Wright, and A. Miller, 556-577. John Wiley & Sons.

[43] Fodor, Jerry. 1974. " Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28, no. 2: 97-115.

[44] Francez, Nasim. 2015. *Proof-Theoretic Semantics*. College Publications.

[45] French, Stephen and James Ladyman. 2003. "Remodeling Structural Realism: Quantum Physics and the Metaphysics of Structure." *Synthese* 136, 31-56.

[46] Gentzen, Gerhard. 1935. "Investigations into Logical Deduction," in *The Collected Papers of Gerhard Gentzen*, ed. M. Szabo, 68-131. Amsterdam: North- Holland. 1969.

[47] Hacking, Ian. 2007. "The Contingencies of Ambiguity." *Analysis* 67, no. 4: 269-277.

[48] Hale, Bob. 2013. *Necessary Beings: An Essay in Ontology, Modality, and the Relations Between Them*. Oxford: Oxford University Press.

[49] Hanks, Peter. 2011. "Structured Propositions as Types." *Mind* 120, no. 377: 11-52.

[50] Hanks, Peter. 2015. *Propositional Content*. Oxford: Oxford University Press.

[51] Hanks, Peter. 2017. " Predication and Rule-Following." In *Philosophy and Logic of Predication*, ed. Piotr Stalmaszczyk, 199-223. New York: Peter Lang.

[52] Haugeland, John. 1998. "Truth and Rule Following." In *Having Thought*, 305-362. Cambridge, MA: Harvard University Press.

[53] Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Malden: Blackwell.

[54] Hintikka, Jaakko. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Ithaca: Cornell University Press.

[55] Hintikka, Jaakko. 1975. *The Intentions of Intentionality and Other New Models for Modalities*. Reidel.

[56] Hlobil, Ulf. 2016. "A Nonmonotonic Sequent Calculus for Inferentialist Expressivists." In *The Logica Yearbook 2015*, ed. P. Arazim and M. Dančák, 87-105. London: College Publications.

[57] Hlobil, Ulf. 2017. "When Structural Principles Hold Merely Locally." In *The Logica Yearbook*, ed. P. Arazim and T Lávika, 53-67. London: College Publications.

[58] Hlobil, Ulf. 2018. "Choosing Your Nonmonotonic Logic: A Shopper's Guide." In *The Logica Yearbook 2017*, ed. P. Arazim and T. Lávička, 109-123. London: College Publications.

[59] Hlobil, Ulf. 2021. "Bilateralist Truth-Maker Semantics for ST and Related Logics." Presentation at UConn Logic Colloquium. Available at: https://www.youtube.com/watch?v=YIf1BWR64Ps&ab_channel=UConnLogicGroup.

[60] Hornsby, Jennifer. 1997. "Truth: The Identity Theory." *Proceedings of the Aristotelian Society* 97: 1-24.

[61] Jubien, Michael. 2009. *Possibility*. Oxford: Oxford University Press.

[62] Kamp, Hans. 1981. "A Theory of Truth and Semantic Representation." In *Formal Methods in the Study of Language*, ed. J. Groenendijk, 277-322. University of Amsterdam.

[63] Kaplan, Daniel. 2018. "A Multi-Succident Sequent Calculus for Logical Expressivists." In *The Logica Yearbook 2017*, ed. P. Arazim and T. Lákiva, 139-154. London: College Publications.

[64] Kearns, Kate. 2011. *Semantics: Second Edition*. New York: Palgrave Macmillan.

[65] Ketonen, Oiva. 1944. *Untersuchungen zum Pradikatenkalkul*, Annales Acad. Sci. Fenn. Ser. A.I. 23. Helsinki.

[66] Kimhi, Irad. 2018. *Thinking and Being.* Cambridge, MA: Harvard University Press.

[67] King, Jeff. 1998. "What Is a Philosophical Analysis?" *Philosophical Studies* 90: 155-179, 1998.

[68] King, Jeff. 2007a. "What in the World Are the Ways Things Might Have Been?" *Philosophical Studies* 133, no. 3 (April): 443-453.

[69] King, Jeff. 2007b. *The Nature and Structure of Content.* Oxford: Oxford University Press.

[70] King, Jeff. 2014. "Naturalized Propositions." In *New Thinking About Propositions*, 47-70. Oxford: Oxford University Press.

[71] King, Jeff. 2018. "W(h)ither Semantics! (?)" *Nous* 54, no. 4: 772-795.

[72] Kratzer, Angelika. 1977. "What 'Can' and 'Must' Can and Must Mean." *Linguistics and Philosophy* 1, no. 3: 337-355.

[73] Kratzer, Angelika. 1986. "Conditionals." *Chicago Linguistics Society* 22, no. 2:1-15.

[74] Kraut, Robert. 1979. "Attitudes and Their Objects." *Journal of Philosophical Logic* 8, no. 2: 197-217.

[75] Kraut, Robert. 1982. "Sensory States and Sensory Objects." *Nous* 16, no. 2: 277-293.

[76] Kraut, Robert. 2010. "Universals, Metaphysical Explanations, and Pragmatism." *The Journal of Philosophy* 107, no. 11: 590-609.

[77] Kremer, Michael. M.S. "The Unity of the Myth of the Given." Presented at *The Wilfrid Sellars Society* in January 2017.

[78] Kripke, Saul. 1959. "A Completeness Theorem in Modal Logic." *The Journal of Symbolic Logic* 24, no. 1: 1-14.

[79] Kripke, Saul. 1965. "Semantical Analysis of Intuitionistic Logic." In *Formal Systems and Recursive Functions*. ed. J. Crossley and M. A. E. Dummett, 92–130. Amsterdam: North Holland Publishing.

[80] Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

[81] Kukla, Rebecca and Mark Lance. 2009. *'Yo!' and 'Lo!'*. Cambridge, MA: Harvard University Press.

[82] Kukla, Rebecca and Mark Lance. 2010. "Perception, Language, and the First-Person." In *Reading Brandom*, ed. J. Wanderer and B. Wiess, 115-128. New York: Routledge.

[83] Kukla, Rebecca and Mark Lance. 2014. "Intersubjectivity and Receptive Experience." *The Southern Journal of Philosophy* 52, no. 1: 22-42.

[84] Ladyman, James and Don Ross. 2007. *Everything Must Go: Metaphysics Naturalized*. Oxford, Oxford University Press.

[85] Lance, Mark. 1996. "Quantification, Substitution, and Conceptual Content." *Nous* 30, no. 4: 481-507.

[86] Lance, Mark and Philip Kremer. 1994. "The Logical Structure of Linguistic Commitment I: Four Systems of Nonrelevant Commitment Entailment." *Journal of Philosophical Logic* 23, no. 4: 369-400.

[87] Lance, Mark and Philip Kremer. 1996. "The Logical Structure of Linguistic Commitment II: Systems of Relevant Commitment Entailment." *Journal of Philosophical Logic* 25, no. 4: 425-449.

[88] Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press.

[89] Lewis, David. 1973. *Counterfactuals*. Basil Blackwell.

[90] Lewis, David. 1975. "Language and Languages." In *Minnesota Studies in the Philosophy of Science*, edited by Keith Gunderson, 3-35. University of Minnesota Press.

[91] Lewis, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8, no. 1: 339-359.

[92] Lewis, David. 1986. *On the Plurality of Worlds*. Wiley Blackwell.

[93] Linnaeus, Carl. 1756. *Systema Naturae, 9th Edition.* Leiden: Theodor Haak.

[94] Löbner, Sebastian. 2002. *Understanding Semantics*. Abington: Routledge.

[95] MacFarlane, John. 2010. "Pragmatism and Inferentialism." In *Reading Brandom*, ed. J. Wanderer and B. Wiess, 81-95. New York: Routledge.

[96] MacFarlane, John. 2021. *Philosophical Logic: A Contemporary Introduction*. New York: Routledge.

[97] Maher, Chaucy. 2012. *The Pittsburgh School of Philosophy: Sellars, Brandom, McDowell.* New York: Routledge.

[98] McDowell, John. 1998. "In Defense of Modesty." In *Meaning, Knowledge, and Reality*. Cambridge, MA: Harvard University Press.

[99] McDowell, John. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.

[100] McDowell, John. 2002. "Reply to Barry Stroud." In *Reading McDowell*, ed. N. H. Smith, 277-279. New York: Routledge.

[101] McDowell, John. 2009. "Avoiding the Myth of the Given." In *Having the World In View*. Cambridge, MA: Harvard University Press.

[102] Mermin, David. "What's Wrong with this Pillow?" *Physics Today* 42, no. 4: 9-11.

[103] Milson, Jared. 2014. "Queries and Assertions in Minimally Discursive Practices." *Proceedings of the Society for the Study of Artificial Intelligence and the Simulation of Behavior.* Goldsmiths Press.

[104] Montague, Richard. 1974. "English as a Formal Language." In *Formal Philosophy: Selected Papers of Richard Montague*, ed. R. Thomason, 188-221. New Haven: Yale.

[105] Negri, Sara and Jan von Plato. 2008. *Structural Proof Theory*. Cambridge: Cambridge University Press.

[106] Nickel, Bernhard. 2013. "Dynamics, Brandom-Style." *Philosophical Studies* 162, no. 2: 333-354.

[107] Okrent, Mark. 2007. *Rational Animals: The Teleological Roots of Intentionality*. Athens, Ohio: Ohio University Press.

[108] Okrent, Mark. 2018. *Nature and Normativity: Biology, Teleology, and Meaning.* New York: Routledge.

[109] O'Leary-Hawthorne, John. 1996. "The Epistemology of Possible Worlds: A Guided Tour." *Philosophical Studies* 84, no. 2/3: 183-202.

[110] Oppenheim, Paul and Hilary Putnam. 1958. *Minnesota Studies in the Philosophy of Science* 2: 3-36.

[111] Orilia, Francesco and Michele Paolini Paoletti. 2020. "Properties." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/properties/

[112] O'Shea, James. 2007. *Wilfrid Sellars: Naturalism With a Normative Turn*. Cambridge: Polity Press.

[113] Partee, Barbara. 1988. "Possible Worlds in Model Theoretic Semantics," in *Possible Worlds in Humanities, Arts, and Sciences*, 93-123. Berlin: De Gruyter.

[114] Partee, Barbara. 2005. "Meaning Postulates and the Lexicon." Lecture 5 notes for *Topics in Formal Semantics*.

[115] Pautz, Adam. 2006. "Can the Physicalist Explain Color Structure in Terms of Color Experience." *Australasian Journal of Philosophy* 84, no. 4/.

[116] Pautz, Adam. 2006b. "Color Eliminitivism." Manuscript.

[117] Pietroski, Paul. 2018. *Conjoining Meanings: Semantics without Truth Values*. Oxford: Oxford University Press.

[118] Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. Palgrave Macmillion.

[119] Plantinga, Alvin. 1976. "Actualism and Possible Worlds." *Theoria* 42: 139-160.

[120] Portner, Paul. 2005. *What Is Meaning?* Blackwell.

[121] Putnam, Hilary. 1975a. "What Is Mathematical Truth." In *Mathematics, Matter, and Method*, 60-78. Cambridge: Cambridge University Press.

[122] Putnam, Hilary. 1975b. "The Meaning of Meaning." *Minnesota Studies in the Philosophy of Science* 7: 131-193.

[123] Quine, W.V.O. 1953. "Two Dogmas of Empiricism." In W.V.O Quine, *From a Logical Point of View*, 20-46. Cambridge, MA: Harvard University Press.

[124] Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

[125] Restall, Greg. 1999. "Negation in Relevant Logics." In *What Is Negation?*, ed. D. M. Gabbay and H. Wansing, 53-76. Dordrecht: Springer.

[126] Restall, Greg. 2005. "Multiple Conclusions." In *Logic, Methodology and Philosophy of Science*, ed. P. Hájek, L. Valdés-Villanueva and D. Westerstahl. College Publications.

[127] Restall, Greg and Francesco Paoli. "The Geometry of Non-Distributive Logics." *The Journal of Symbolic Logic* 70, no. 4: 1108-1126.

[128] Ripley, David. 2013. "Paradoxes and Failures of Cut." *Australasian Journal of Philosophy* 91, no. 1: 139-164.

[129] Rumfitt, Ian. 2000. "Yes and No." *Mind*, 109 (436), 781-823.

[130] Rumfit, Ian. 2008. "Knowledge by Deduction. *Grazer Philosophische Studien* 77: 61-84.

[131] Sampson, William. 1819. *Is a Whale a Fish?: an accurate report of the case of James Maurice against Samuel Judd, tried in the Mayor's Court of the city of New-York, on the 30th and 31st of December 1818: wherein the above problem is discussed theologically, scholastically, and historically.* New York: C.S. Van Winkle.

[132] Sellars, Wilfrid. 1953. "Inference and Meaning." *Mind* 62, no. 247: 313-338.

[133] Sellars, Wilfrid. 1954. "Some Reflections on Language Games." *Philosophy of Science* 21, no. 3: 204-228.

[134] Sellars, Wilfrid. 1956. " Empiricism and the Philosophy of Mind." *Minnesota Studies in the Philosophy of Science* 1:253-329.

[135] Sellars, Wilfrid. 1962. "Philosophy and the Scientific Image of Man." In *Science, Perception, and Reality*. Atascadero: Ridgeview Publishing.

[136] Sellars, Wilfrid. 1963. "Abstract Entities." *The Review of Metaphysics* 16, no. 4: 627-671.

[137] Sellars, Wilfrid. 1968. *Science and Metaphysics*. London: Routledge & Keegan Paul.

[138] Sellars, Wilfrid. 1969. "Language as Thought and as Communication." *Philosophy and Phenomenological Research* 29, no. 4: 506-527.

[139] Sellars, Wilfrid. 1974. "Meaning as Functional Classification." *Synthese* 27, no. 3-4: 417-437.

[140] Sellars, Wilfrid. 1979. *Naturalism and Ontology*. Atascadero: Ridgeview Publishing.

[141] Shapiro, Lionel. 2018. "Logical Expressivism and Logical Relations." In *From Rules to Meanings: New Essays on Inferentialism*, ed. O. Beran, V. Kolman, and L. Koren, 179-195. New York: Routledge.

[142] Shimamura, Shuhei. 2017. "A Nonmonotonic Modal Relevant Sequent Calculus." In *Logic, Rationality, and Interaction LORI 2017*. Lecture Notes in Computer Science, vol 10455, ed. A. Baltag, J. Seligman, and T. Yamada, 570-584. Berlin: Springer.

[143] Shimamura, Shuhei. 2019. "A First-Order Sequent Calculus for Logical Inferentialists and Expressivists." In *Logica Yearbook 2018*, ed. I. Sedlá and M. Blicha, 211-228. London: College Publications.

[144] Simonelli, Ryan. 2020. "The Normative/Agentive Correspondence." *Journal of Transcendental Philosophy*.

[145] Simonelli, Ryan. 2021. "Sellars's Ontological Nominalism." *European Journal of Philosophy*.

[146] Simonelli, Ryan. Forthcoming. "Sellars's Two Worlds." In *Reading Kant with Sellars*, ed. L. C. Seiberth and M. Ranee. Routledge.

[147] Simonelli, Ryan. M.S.a. "Considering the Exceptions: On the Failure of Cumulative Transitivity for Indicative Conditionals." Available at ryansimonelli.com/papers.

[148] Simonelli, Ryan. M.S.b. "Bringing Bilateralisms Together (And Pulling them Apart Again)." Available at ryansimonelli.com/papers.

[149] Simonelli, Ryan. M.S.c. "Why Must Incompatibility Be Symmetric?" Available at ryansimonelli.com/papers.

[150] Simonelli, Ryan. M.S.d. "Simply Substructural Semantics." Available at ryansimonelli.com/papers.

[151] Simonelli, Ryan. M.S.e. "How to Be a Hyper-Inferentialist." Available at ryansimonelli.com/papers.

[152] Simonelli, Ryan. M.S.f. "Propositions and the Power to Represent." Available at ryansimonelli.com/papers.

[153] Smiley, Timothy. 1996. "Rejection." *Analysis* 56, no. 1: 1-9.

[154] Soames, Scott. 2010. *Philosophy of Language*. Princeton: Princeton University Press.

[155] Soames, Scott. 2014. "Cognitive Propositions." In *New Thinking About Propositions*, 91-124. Oxford: Oxford University Press.

[156] Soames, Scott. 2015a. *Rethinking Language, Mind, and Meaning.* Princeton: Princeton University Press.

[157] Soames, Scott. 2015b. "Reply to Peter Hanks and Ray Buchanon." APA Session on *New Thinking about Propositions*, St. Lous Missouri.

[158] Soames, Scott. 2019. *The World Philosophy Made.* Princeton: Princeton University Press.

[159] Stalnaker, Robert. 1978. "Assertion." *Syntax and Semantics* 9: 315-332. Reprinted in Stalnaker, Robert. 1999. *Context and Content*, 78-95.

[160] Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

[161] Stanley, Jason. 2006. "The Use Theory of Meaning." *Leiter Reports: A Philosophy Blog*. https://leiterreports.typepad.com/blog/2006/03/the_use_theory_.html.

[162] Stovall, Preston. 2021. "Essence as Modality: A Proof-Theoretic and Nominalist Analysis." *Philosopher's Imprint* 21, no. 7: 1-28.

[163] Stovall, Preston. 2022. *The Single Minded Animal: Shared Intentionality, Normativity, and the Foundations of Discursive Cognition*. New York: Routledge.

[164] Stroud, Barry. 2002. "Wittgenstein's 'Treatment' Of the Quest for 'A Language Which Describes My Inner Experiences and Which Only I Myself Can Understand."' In *Meaning, Understanding, Practice*, 67-80. Oxford: Oxford University Press.

[165] Stroud, Barry. 2018. "Seeing What is So." In *Seeing, Knowing, and Understanding: Philosophical Essays*, 86-100. Oxford: Oxford University Press.

[166] Szabó, Zoltán Gendler. 2019. "Semantic Explanations." In *Oxford Studies in the Philosophy of Language*, ed. E. Lepore and D. Sosa, 240-275. Oxford: Oxford University Press.

[167] Tanter, Kai. 2021. "Subatomic Inferences: an Inferentialist Semantics for Atomics, Predicates, and Names." *Review of Symbolic Logic.*

[168] Thomason, Richard. 1974. "Introduction." In *Formal Philosophy: The Selected Papers of Richard Montague*, edited by Richard Thomason, 1-69. New Haven: Yale University Press.

[169] Thomasson, Amie. 2020. *Norms and Necessity*. Oxford: Oxford University Press.

[170] Tomasselo, Michael. 2014. *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

[171] Van Inwagen, Peter. "Properties." In *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, ed. T. Crisp, M. Davidson, D. Vander Laan, 15-34. Dordrecht: Springer.

[172] Veltman, Frank. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic* 25, 221-261.

[173] Whiting, Daniel. 2006. "Conceptual Role Semantics." *Internet Encyclopedia of Philosophy*.

[174] Willer, Malte. 2013. "Dynamics of Epistemic Modality." *Philosophical Review* 122, no. 1: 45-92.

[175] Willer, Malte. 2021. "Two Puzzles about Ability *Can*." *Linguistics and Philosophy* 44, no. 3: 551-586.

[176] Wittgenstein, Ludwig. 1953/1958. *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford: Basil Blackwell.

[177] Yalcin, Seth. 2014. "Semantics and Metasemantics in the Context of Generative Grammar." In *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman, 17-54. Oxford: Oxford University Press.

[178] Yalcin, Seth. 2018. "Semantics as Model-Based Science." In *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, edited by Derek Ball and Brian Rabern, 334-360. Oxford: Oxford University Press.