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 sentences. 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 white, it cannot be black," which ostensibly express *worldly* and specifically *metaphysical* knowledge, as really functioning to express the scorekeeping principles that determine the discursive significance of the sentences of the form "x is white" and "x is black." This precisely spells out a version of what has been called "modal normativism," a position originally charted out by Sellars and developed and defended most recently by Amie Thomasson (2020). This modal normativist account of conditionals of the above sort will enable 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 black is incompatible with the property of being white in the sense that it's not possible for something to instantiate the property of being black and also instantiate the property of being white, or, to put it differently, if something instantiates the property of being black, 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 "black," "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 are 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 is what I've argued really. 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 it out modal normativism, a (metaphysically) modalized sentence of the form "Necessarily φ" is true just in case φ is an object-langauge 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 with the 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 form $\varphi \rightarrow \psi$ expresses that

[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

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 $\varphi \to \neg\psi$, 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. For, while Thomasson claims that distinctively *modal* properties, such as the property of being necessarily incompatible with 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

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 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 of 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 the main reason why Brandom has stopped using the notion of incompatibility entailment in recent work.

doesn't pressupose 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 Kremer and Lance (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 expressivsm that has been taken by Brandom and his collaborators in recent years.

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), members of this group have formally developed a conception of expressivism originally put forward by Brandom (2008), putting forward a general program for logical expressivsim 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. I'm getting off the train, however,

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

when it comes to conditionals, or, at least, when it comes to the conditionals that are particularly pertinent to the project here.

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 langauge to a logically complex language $\mathcal{L}$ that includes sentences containing, for instance, the expressions "→" and "∧," which related by an *extended consequence* relation $\vdash$. To see how such a calculus can be thought of as an "expressivist logic," suppose our base consequence relation $\vdash_0$, relating sentences of $\mathcal{L}_0$ contains the following:

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

So, speakers of $\mathcal{L}_0$ infer "*a* is darker than *b*" from both "*a* is a 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 "→" and "∧." 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 et al. :

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

Unlike speakers of $\mathcal{L}_0$, speakers of $\mathcal{L}$ *do* have a way of making the inferential relation that obtains between "*a* is a 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:

$$\cfrac{\cfrac{\text{“}a\text{ is white,” “}b\text{ is gray”} \vdash \text{“}a\text{ is darker than }b\text{”}}{\text{“}a\text{ is white and }b\text{ is gray”} \vdash \text{“}a\text{ is darker than }b\text{”}}\,L\wedge}{\vdash \text{“If }a\text{ is white and }b\text{ is gray, then }a\text{ is darker than }b\text{”}}$$

Since this sentence follows from the empty set of sentences, it might be though 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:

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

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 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. To show how deep the issue here is, however, let us focus on the positive conditional rule, which Brandom (2018), though a explicitly 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, this would amount to taking anyone who we take to be committed to q

to be 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, by thinking of conditionals as expressing scorekeeping principles makes sense of why it does. 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 the has been laid out. 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 $\vdash \oplus_\alpha\langle\varphi \rightarrow \psi\rangle$ 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 $\vdash \oplus_\alpha\langle\varphi \rightarrow \psi\rangle$, 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 the failure of the deduction theorem, on this framework, is the result of the cross-perspectival interplay of "$\vdash$" and "$\rightarrow$". In the context of a scorekeeping principle that we hold, "$\vdash$" is the location we use to think of *our own* scorekeeping principles, whereas "$\rightarrow$" is the locution we use to to think about the scorekeeping principles of *other scorekeepers*. In an expression of the from '$\varphi \vdash \psi \rightarrow \chi$," the "$\vdash$" is expressing *our* scorekeeping principle, whereas the "$\rightarrow$" is expressing the scorekeeping principle of an arbitrary *other scorekeeper*. The reason why the move from $\Gamma, \varphi \vdash \psi$ to $\Gamma \vdash \varphi \rightarrow \psi$ 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. 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

[3]The account of objectivity developed in the next chapter, which has its roots in Brandom's later work, will ultimatley have to be merged with the perspectival account of objectivity for a full account.

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. Logical vocabulary has thus far been modeled in terms of our principles for *attributing* normative statuses to someone else when *they* utter some sentence. Our expressivist account of conditionals, however, should articulate that we grasp when *we* are prepared to *undertake* a commitment to a conditional sentence *ourselves*, given how we attribute commitments to others. 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 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. Thus, while we *can* model the function of conditionals from the perspective of the *hearer* of conditional utterances, defining updates for them in the way we've done in the last chapter (see Appendix for the first step towards a system), insofar as we're expressivists about the use of conditionals, there is reason to look for a rather different kind of system, one that articulates the perspective of the potential *speaker* of conditional utterances. Let us now turn to such a system.

5.4 Lance and Kremer's Commitment Logic

The way of introducing conditionals I will propose is owed to Lance and 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 Kremer and Lance 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	$\mid\ \oplus_{\alpha_1}\langle p\rangle$	asm.
2	$\mid\ \oplus_{\alpha_1}\langle q\rangle$	PA $(1, \oplus_{\alpha}\langle p\rangle \vdash \oplus_{\alpha}\langle q\rangle)$
3	$p \rightarrow q$	$\rightarrow_I (1, 2)$

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]

[4]Though Kremer and Lance 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,

> 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 "$\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$." An embedded conditional of the form $(\varphi \rightarrow (\psi \rightarrow \chi))$ 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:

> $\rightarrow_I$: 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 $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n}\langle\varphi \rightarrow \psi\rangle$, where $n \geq 0$.

Given this introduction rule, there is a natural corresponding elimination rule, a form of modus ponens, an instance of which is shown below:

1	$\oplus_{\alpha_1}\langle(\varphi \rightarrow \psi)\rangle$	asm.
2	$\quad\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$	asm.
3	$\quad\oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\psi\rangle\rangle$	$\rightarrow_E$ (1, 2).

The thought here is that if α_1 is committed to $(\varphi \rightarrow \psi)$, then α_1 will score anyone they score as committed to φ to be committed to ψ. So, from $\oplus_{\alpha_1}\langle\varphi \rightarrow \psi\rangle$ 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:

---

they never actually provide a way of integrating such material relations of committive consequence. That is what this rule does.

$\to_E$: From $\oplus_{\alpha_1} \ldots \oplus_{\alpha_n} \langle \varphi \to \psi \rangle$ 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 socorekeeper is committed to a conditional of the form $\varphi \to \psi$, 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* consquences. 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 $((\varphi \to \psi) \wedge (\psi \to \chi)) \to (\varphi \to \chi)$, 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	$\mid\ \oplus_{\alpha_1}\langle((\varphi \to \psi) \wedge (\psi \to \chi))\rangle$	asm.
2	$\mid\ \oplus_{\alpha_1}\langle((\varphi \to \psi))\rangle$	$\wedge_{E_L}$ (1)
3	$\mid\ \oplus_{\alpha_1}\langle((\psi \to \chi))\rangle$	$\wedge_{E_R}$ (1)
4	$\mid\ \mid\ \oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\varphi\rangle\rangle$	asm.
5	$\mid\ \mid\ \oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\psi\rangle\rangle$	$\to_E$ (2, 4)
6	$\mid\ \mid\ \oplus_{\alpha_1}\langle\oplus_{\alpha_2}\langle\chi\rangle\rangle$	$\to_E$ (3, 5)
7	$\mid\ \oplus_{\alpha_1}\langle((\varphi \to \chi))\rangle$	$\to_I$ (4-6)
8	$((\varphi \to \psi) \wedge (\psi \to \chi)) \to (\varphi \to \chi)$	$\to_I$ (1-7)

It's easy to construct similar proofs of all of the structural principles enforced by the update semantics, as shown in the previous chapter, such as Monototnicity, Exchange, 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.[5]

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, $p \to (q \to p)$ is not provable in this system. The following "proof" fails to accord with the conditional introduction rule:

1	$\quad \mid \oplus_{\alpha_1}\langle p \rangle$	asm.
2	$\quad \mid \quad \mid \oplus_{\alpha_1}\langle \oplus_{\alpha_2}\langle q \rangle\rangle$	asm.
3	$\quad \mid \quad \mid \oplus_{\alpha_1}\langle p \rangle$	reit. (1)
4	$\quad \mid \oplus_{\alpha_1}\langle q \to p \rangle$	$\to_I$ (2 – 3)? (fallacious step)
5	$p \to (q \to p)$	$\to_I$ (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 $\oplus_{\alpha_1}\langle (q \to p)\rangle$. 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, $p \to (q \to p)$ 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 considered in the context of the ROLE approach, in which $p \to (q \to p)$ straightforwardly follows from CO and the Deduction Theorem.

[5]In order to be able to express (instances of) *bilateral* scorekeeping principles, such as Reversal or Bilateral Reducto, 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 in the Appendix, 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.

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:

1	$\mid\ \oplus_{\alpha_1}\langle p \wedge q\rangle$	asm.
2	$\mid\ \oplus_{\alpha_1}\langle p\rangle$	$\wedge_{E_L}$ (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.[6]

[6]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 commitive 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.

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 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 Kremer and lance 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 Kremer and Lance 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 (See 1994, 383). Though Kremer and Lance 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 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 Kremer and Lance call C1.

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:[7]

$$
\begin{array}{lll}
1 & \quad \boxed{a} & \\
2 & \quad \vdots & \\
3 & \quad \Phi(a) & \\
4 & \forall x(\Phi(x)) & \forall_I
\end{array}
$$

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:

$$
\begin{array}{lll}
1 & \quad \boxed{a} & \\
2 & \quad\quad \oplus_{\alpha_1}\langle Fa \rangle & \text{asm.} \\
3 & \quad\quad \vdots & \\
4 & \quad\quad \oplus_{\alpha_1}\langle Ga \rangle & \\
5 & \quad Fa \rightarrow Ga & \rightarrow_I \\
6 & (\forall x)(Fx \rightarrow Gx) & \forall_I
\end{array}
$$

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

[7] 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.

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

$\oplus_\alpha\langle a \text{ is white}\rangle \vdash \oplus_\alpha\langle \text{It's not the case that } a \text{ is black}\rangle$

Is conceived of as generated from the following scorekeeping principle:

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

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

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

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 principles on sentences in which *any* singular term is substituted in for the variable. Thus, we can reason as follows:

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

Thus, because we score an arbitrary agent who we take to be committed to "*a* is white," for an arbitrary name *a*, to be precluded from being entitled to "*a* is black" and so committed to "It's not the case that *a* is black," we can assert, for an arbitrary *a*, "If *a* is white, then it's not the case that *a* is black," and that enables us to assert "For all *x*, if *x* is white, then it's not the case that *x* is black." 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, $(\forall x)(x$ is black$) \rightarrow ((\forall y)(\neg y$ is darker than $x))$. We saw in the last chapter that we have following material scorekeeping principle (on frames):

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

With this scorekeeping principle, we can reason as follows:

1	$\boxed{a}$	
2	$\oplus_{\alpha_1}\langle a \text{ is black}\rangle$	asm.
3	$\boxed{b}$	
4	$\oplus_{\alpha_1}\langle a \text{ is black}\rangle$	reit.
5	$\oplus_{\alpha_1}\langle \neg b \text{ is darker than } a\rangle$	PA (1, $\oplus\langle Bx\rangle \vdash \ominus_{\alpha}\langle Dyx\rangle$)
6	$\oplus_{\alpha_1}\langle \forall y(\neg y \text{ is darker than } a)\rangle$	$\forall_I$ (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	$\oplus_{\alpha_1}\langle a \text{ is the same shade as } a\rangle$	($\vdash \oplus_{\alpha}\langle Sxx\rangle$).
3	$a \text{ is the same shade as } a$	$\oplus_E$
4	$\forall x(x \text{ is the same shade as } x)$	$\forall_I$

Note that this system is one that functions to provide object-langauge 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 derrive, 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:[8]

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

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:

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

Now, given the PA rule, we can derive the universally quantified conditional used in this definition of the property of being black as follows:

[8] 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: $\forall x((Bx \rightarrow \forall y((Gy \vee Wy) \supset Dxy)$

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

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

$$\oplus\langle Wy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle$$
$$\oplus\langle Gy\rangle, \oplus\langle Bx\rangle \vdash \oplus\langle Dxy\rangle$$

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:

$$\oplus\langle Bx\rangle \vdash \ominus_{\alpha}\langle Dyx\rangle$$

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 worldly 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) articualtes 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 "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 langauge, have everything to do with language. Moreover, not only do they have everything to do with semantics, 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 p such that σ contains both $\oplus_{\alpha_1}\langle p\rangle$ and $\ominus_{\alpha_1}\langle p\rangle$.

The new non-primitivist actualist conception of possible worlds (e.g. 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 scoreards. To say "The world could be such that a is black, and it could be such that a is white, but it can't be such that a is both black and white," is a way of expressing, in worldly vocabulary, that there is a coherent set of normative assignments that contains $\oplus_{\alpha_1}\langle a \text{ is black}\rangle$ 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 $\oplus_{\alpha_1}\langle a \text{ is black}\rangle$ and $\oplus_{\alpha_1}\langle a \text{ is white}\rangle$, since commitment to "$a$ is black" precludes entitlement to "a 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 black 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 that that the reason no coherent scorecard contains $\oplus\langle a \text{ is white}\rangle$ and $\oplus\langle a \text{ is black}\rangle$ is that these are positions of the form $\oplus\langle x \text{ is white}\rangle$ and $\oplus\langle x \text{ is black}\rangle$, and score is kept in accordance with the principle $\oplus_\alpha\langle x \text{ is white}\rangle \vdash \ominus_\alpha\langle x \text{ is black}\rangle$

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 ambitious 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 $\oplus\langle p\rangle \in \sigma$ and *false* if $\ominus\langle p\rangle \in \sigma$, and it conforms to the coherence requirement since there is no sentence p such that $\oplus\langle p\rangle \in \sigma$ and $\ominus\langle p\rangle \in \sigma$.

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. 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:

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

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

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

And requiring coherence, given this scorekeeping principle, puts the very same constraint on scorecards that the above postulate puts on worlds, ruling out any scorcared in which there is some term τ such $\oplus_{\alpha_1}\langle Wx\rangle \in \sigma$ and $\oplus_{\alpha_1}\langle Bx\rangle \in \sigma$. 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 clear how our reconstruction of worlds in terms of scorecards explains why, for instance, there is no world in which both "a is black" 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, is supposed to (in part) explain the fact that "a is black" 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 $\ominus_{\alpha_1}\langle p\rangle \in \sigma$, 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 $\oplus_\alpha\langle\varphi\rangle \in \sigma$ just in case $\oplus\langle\varphi\rangle \in \sigma$ and $\oplus\langle\psi\rangle \in \sigma$, and $\ominus_\alpha\langle\varphi \wedge \psi\rangle \in \sigma$ just in case $\ominus\langle\varphi\rangle \in \sigma$ or $\ominus\langle\psi\rangle \in \sigma$. 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, 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. All of these basic "worldly" notions 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. One might move from the fact that extra-worldly semantic models are *predictive* to the fact 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 and negation along with 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 $\neg\varphi$. Our reconstruction of possible worlds in scorekeeping terms, however, makes clear that there is a *reflection* of this fact here, not any explanation of it.

The transposition of these fact about possible worlds into this normative vocabulary is that is 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 $\oplus\langle\neg\varphi\rangle$ will in fact be the complement of those containing $\oplus\langle\varphi\rangle$, 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 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 $\oplus\langle\varphi\rangle$, will, given the application of scorekeeping principles to it, 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 obtaining 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 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 they're their committed to and precluded from being entitled to is updated in the following way:[9]

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,

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, the we can 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

[9] Unlike Stalnaker's proposal, the context states that we are treating as updated in thi way are

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 one can genuinely explain the worldly contents appealed to in a possible worlds semantics by starting with a normative pragmatic account of discursive roles (that's just what I have done in this chapter), the explanaion 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 *expain* 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 mere *reflections* of semantic competence.[10]

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

[10] One way of hearing this claim, putting it into is that purpose of discursive role semantics is that it is able to provide a *metasemantic* theory that underlies. Recent proponents of inferentialism (Murzi and St) as well as Recently, theorists who've been drawn to modal normativism

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, and considering , 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 one's that we've explicitly considered here, and so now have a very clear sense of how to understand. 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, 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 conception 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.[11] 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 elsewhere (Simonelli, M.S.), 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 one, 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 adquate 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:

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

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

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

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

[11] Consider the proposal of Kukla and Lance (2009), drawing on

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

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

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, 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 on the proper articuation of the notion of *seeing*, conceived of as an actualization of a perceptual capacity.[12] 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 methedology, for the metaphysical structure at issue here can be understood as a reification of scorekeeping principles involving the

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

terms "sees."[13] 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 $\langle x$ is red$\rangle$" and committed to "x is red" oneself, but also committed to "n is committed to $\langle x$ is red$\rangle$." 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, Japanese, and various other natural langauges 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's 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

[13]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).

phrase "The 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," acheived here by italicization. On the account offered here, which is owed to Sellars (), rather than functioning to pick out some non-linguistic thing, the italizied "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, clearly, 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.[14]

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. 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. The lack of correspondence between the properties grasped by the English speaker and the properties grasped by

[14] 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.

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.