1

Introduction

1.1 Introduction

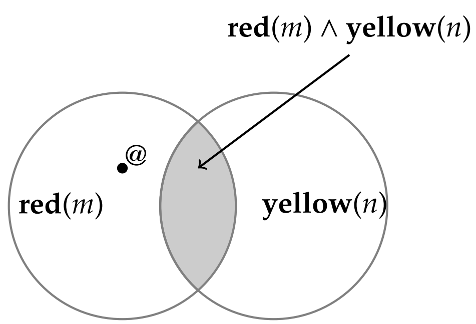
This book is an introduction to the formal tools used to pursue an *inferentialist* or *proof-theoretic* approach to linguistic meaning. Such an approach stands in opposition to what is currently the dominant approach to meaning in philosophy and linguistics: a *representationalist* or *model-theoretic* approach. I'll use the terms "inferentialist" and "proof-theoretic" (and, correspondingly "representationalist" and "model-theoretic") to refer, respectively, to the guiding philosophical idea and the main technical approach for formally implementing that idea. This book is primarily a logic book, aimed at introducing you to the formal tools so that you can technically pursue inferentialist semantics yourself, but these formal tools have been developed in pursuit of a philosophical program. This chapter introduces that basic program.

1.2 Representationalist, Model-Theoretic Approaches

Let's start with representationalism. According to representationalism, the meaning of a sentence is to be understood in terms of the way it represents the world as being. For instance, "Mars is red" represents the world as being such that Mars is red. We might think of this in the following terms: there is a whole bunch of possible ways that the world can be, a whole bunch of "possible worlds," as it were. There is, for instance, a possible world in which Mars is yellow (perhaps its surface is covered with sulfur in that world), there's a possible world in which Mars is red but is three times its actual size, there's a possible world in which the positions of Mars and Jupiter are swapped, and so on. However things may possibly be, there's a possible world in which things are that way. The meaning of the sentence "Mars is red," the way it represents things as being, carves out a specific subset of those worlds—all of those worlds in which Mars is red. This thought forms the basis for the standard representationalist approach in formal semantics that is dominant today.

A formal semantic theory aims to assign *semantic values* to sentences. The semantic value of a sentence is a mathematical construction that serves as the formal model of the meaning of that sentence. The basic thought behind assigning such semantic values to sentences is that meanings have certain properties, and we want a mathematically precise account of how it is that they have the properties that they do. Most notably, it's widely thought that meanings are *compositional*. That is, the meaning of a complex sentence is determined by the meanings of its parts. For instance, the meaning of "Mars is red and Neptune is yellow" is determined by the meanings of "Mars is red" and "Neptune is yellow." A formal semantic theory, insofar as its compositional, will make this thought precise by assigning semantic values to "Mars is red" and "Neptune is yellow," and assigning an operation to "and," such that the semantic value of "Mars is red and Neptune is yellow" can be computed from the semantic values of its parts. Assigning sets of possible worlds to sentences as their semantic values gives us a straightforward way of doing this.

The set of possible worlds assigned to "Mars is red" are those in which Mars is red, and the set of possible worlds assigned to "Neptune is yellow" are those in which Neptune is yellow. What set of possible worlds will be assigned to "Mars is red and Neptune is yellow"? Well, this will just be the *intersection* of the worlds assigned to "Mars is red" and those assigned to "Neptune is yellow." We might depict this as follows:



Here, the circle on the left represents the set of worlds in which "Mars is red" is true, and the circle on the right represents the set of worlds in which "Neptune is yellow" is true. The shaded *intersection* of these two sets—the set consisting of the worlds that are elements of both sets—is the set of worlds in which "Mars is red and Neptune is yellow" is true. Since the actual world (which I've depicted with a dot and labeled "@") is not in the shaded region, the sentence is false.

Let us lay this approach out a bit more officially. Suppose we are given a language $\mathcal{L}$, containing names such as "Mars" and "Neptune," 1-place predicates such as "is red" and "is yellow," and logical operators such as "and," "or," and "not." Given how expressions of these different types can combine syntactically, the language contains an infinite set of logically complex sentences. Our job, as semantic theorists, is to assign semantic values to the basic expressions of the language, along with rules for how they combine, so as to assign semantic values for the infinite set of sentences the language contains. We know what kind of semantic values we want for complete sentences like "Mars is red," namely, sets of worlds, and we know which sets of worlds we want: "Mars is red" should be assigned the set of worlds in which Mars is red. So we want to assign semantic values to words like "red" and "Mars" so we get this result. There are a few different technically equivalent ways of doing this; I'll lay out one standard one.

Intuitively, to know the meaning of "red" is to be able to take any possible world and single out the set of red things in that world, whatever they happen to be. Accordingly, we might formally model the meaning of "red" as a function that maps each possible world to the set of red things in that world. Officially, we suppose we have a domain D of all of the objects, across all of the possible worlds. The powerset of D, $\mathcal{P}(D)$, is the set of all sets of elements of D. The semantic value of a predicate like "red" will a function that maps each world in W to some set in $\mathcal{P}(D)$, in this case, the set of red things in that world. Officially, we can state it like this:

$$[\![\textbf{red}]\!] = f : W \to \mathcal{P}(D)$$

$$[\![\textbf{red}]\!](w) = \{x \mid x \text{ is red in } w\}$$

Likewise, the meaning of "yellow" maps each world to the set of yellow things in that world:

$$[\![\textbf{yellow}]\!] = f : W \to \mathcal{P}(D)$$

$$[\![\textbf{yellow}]\!](w) = \{x \mid x \text{ is yellow in } w\}$$

We might suppose that the semantic value of a proper name like "Mars" or "Neptune," is just the object named by that name. So,

$$[\![m]\!] = \text{Mars}$$

$$[\![n]\!] = \text{Neptune}$$

With semantic values of these types defined, we can state a *composition rule* that tells us how to compute the semantic value of a complete sentence consisting in a 1-place predicate concatenated with a name. In particular, if we have a 1-place predicate $\mathbf{f}$ and a name a, the set of worlds assigned to the simple predicate sentence $\mathbf{f}(a)$ is defined as follows:

$$[\![\mathbf{f}(a)]\!] = \{w \mid [\![a]\!] \in [\![P]\!](w)\}$$

Thus, considering the particular application of this rule that yields the semantic value of "Mars is red," we have the following:

$$[\![\textbf{red}(m)]\!] = \{w \mid [\![m]\!] \in [\![red]\!](w)\} = \{w \mid \text{Mars is among the red things in } w\}$$

Likewise, for "Neptune is yellow." In this way, we assign semantic values for "red" and "yellow," "Mars" and "Neptune" so that we get semantic values for "Mars is red" and "Neptune is yellow" that codify the circumstances in which these sentences are true. "Mars is red" is true in a world just in case Mars is among the red things in that world, and "Neptune is yellow" is true in a world just in case Neptune is among the yellow things in that world. Since the actual world is one in which Mars is among the red things and Neptune is not among the yellow things, "Mars is red" is true and "Neptune is yellow" is false.

In this way, we are able to assign semantic values to atomic sentences on the basis of the semantic values assigned to names and predicates. Now consider the assignment of semantic values for logically complex sentences such as "Mars is red and Neptune is yellow" or "Mars is not red or Neptune is not yellow." Having assigned simple sentences sets of possible worlds as semantic values, it is straightforward to define We can now recursively define semantic values for atomic sentences, negations, conjunctions, and disjunctions as follows:

$$\mathbf{S}_{\neg} : [\![\neg A]\!] = W - [\![A]\!]$$
$$\mathbf{S}_{\wedge} : [\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!]$$
$$\mathbf{S}_{\vee} : [\![A \wedge B]\!] = [\![A]\!] \cup [\![B]\!]$$

So, the set of worlds assigned to $\neg A$ is the set of worlds that are not elements of $[\![A]\!]$, the set of worlds assigned to $A \wedge B$ is the set of worlds that are elements of both $[\![A]\!]$ and $[\![B]\!]$, and the set of worlds assigned to $A \vee B$ is the set of worlds that are elements of either $[\![A]\!]$ or $[\![B]\!]$. This gives us just the truth-conditions for complex sentences that we want. For instance, "Mars is red and Neptune is yellow" is false since the actual world, though an element of the set of worlds assigned to "Mars is red," is not an element of the set of worlds assigned to "Neptune is yellow."

This is the basis of the sort of semantic theory that is actively being developed by linguistics and philosophers of language. The main project of semantics, as it's currently practiced by the vast majority linguists and philosophers of language who call themselves "semanticists," takes something like this basic representationalist picture as its starting point. Indeed, "semantics," as the term is currently used in linguistic and philosophy of language, implicitly means "representationalist, model-theoretic semantics."[1] This book pursues a different kind of semantics.

Historical Note:

Representationalist, model-theoretic semantics of this sort owes its origins to such figures as Alfred Tarski, Rudolf Carnap, and Richard Montague. Extensional, model-theoretic semantics is originally owed largely to Tarski, who showed how one can systematically assign truth-

[1]For a very good more in-depth introduction to representationalist model-theoretic semantics, developed in this sort of possible worlds framework, see [3].

conditions to sentences of a formal language such as first-order predicate logic. Carnap was pioneering in the development of possible worlds semantics, using the idea of possible worlds (or, as he called them "state descriptions") to define such notions as possibility and necessity. This marked the transition from what is called *extensional* semantics, where a predicate is just assigned the set of things to which it applies, to *intensional* semantics, where a predicate is assigned a function from possible worlds to extensions. With this transition, the basic model-theoretic framework became much more powerful.

Montague brought these ideas together and extended them to apply to natural languages such as English. Montague's basic approach proceeded on his proclamation, "I reject the contention that an important theoretical difference exists between formal and natural languages." Thus, the semantics for English expressions could be undertaken in just the way that a formal semantics. To account for the complexity of the grammar of a natural language such as English, he used a logical type system to account for different kinds of expressions such as adverbs. In this way, formal semantics for natural language was born.

Since Montague, model-theoretic semantics has been largely pursued by linguists, extending the approach to apply to all different kinds of expressions of natural language. Some notable figures include Barbara Partee, Irene Heim, and Angelika Kratzer, who each were hugely influential in the development of formal semantic theory capable of accounting for the complexities of a natural language such as English. Some features of natural language addressed by these theorists include *nominalization* (accounting, for instance, for the transition from "The ball is red," where "red" functions as an adjective, to "Red is my favorite color," where "red" functions as a noun), *anaphora* (accounting, for instance, for the relation between "A dog walked into the room" and "It sat down," where "It" refers to the previously mentioned dog), and *modality* (accounting, for instance, for the different but related senses of "may" in "You may go to the bathroom" and "It may rain later").

It is worth being clear that, while this book pursues an alternative to representationalist, model-theoretic semantics, ultimately, the two approaches should not be seen as *competing*, but as *complementary*. Exactly how to unify the two approaches and understand their relation is a big task that is no the topic of this book. It's worth emphasizing, however, that inferentialists do not need to start from scratch. Rather, they can and should draw on the work of representationalist model-theoretic semanticists when possible. It seems to me that many of the core ideas developed by these authors, though originally put forward in a representationalist, model-theoretic setting, can be adapted to an inferentialist setting. Thus, even if one is interested in developing an inferentialist approach to meaning, one would be well-advised to be well-read in representationalist, model-theoretic semantics.

1.3 Modesty and Full-Bloodedness

Look back at the semantic clauses for "red" and "yellow" we provided in the previous section, and note that, in specifying the meaning of these words in terms of the things to which they apply in any possible world, we *use* these very words in specifying their meanings. A semantic theory of this sort treats the meanings of all grammatically similar predicates exactly the same, using each such predicate to specify the function that is identified as the semantic value of the predicate. An interesting fact about a semantic theory of this sort is that one doesn't actually need to *know* the meanings of expressions for which one is constructing this sort of semantic theory in order to construct the semantic theory; one simply needs to know their grammatical classification. Suppose, for instance, you tell me that the language contains the term "bloog," which is an adjective (the same grammatical category as "red and "yellow"). Even without having the slightest clue what this expression means, I know just how to state the semantic entry that specifies its meaning:

$$[\![\textbf{bloog}]\!] = f : W \to \mathcal{P}(D)$$

$$[\![\textbf{bloog}]\!](w) = \{x \mid x \text{ is bloog in } w\}$$

Of course, such an entry would be of absolutely no help in explicating the meaning of "bloog" to someone who did not already grasp the meaning of this term. It does not, we might say, provide a *substantive account* of the meaning of "bloog" That is, it doesn't say *what it is* for something to be bloog. Likewise, the semantic entries for "red" and "yellow" provided in the previous section do not provide a substantive account of the meanings of "red" or "yellow," in that they do not say what it is for something to be red or what it is for something to be yellow. That fact will not come as a surprise to semantic theorists who develop representationalist, model-theoretic semantic theories: giving an account of the what it is for something to be red or yellow—accounting for the the *conceptual contents* expressed by the expressions "red" and "yellow"—is simply not the project in which they are engaged.

Standard representationalist semantic theories, developed in model-theoretic frameworks like the one sketched above, are what Michael Dummett calls *modest* semantic theories. They do not aspire to give an account of the concepts expressed by the expressions of the language for which they are giving a semantic theory. Accordingly, it is completely permissible to *use* the word "red" in order to specify the condition under which it is true of something. If pitched as an account of the concept expressed by "red," such a clause would be obviously circular, appealing to the very concepts for which it is supposed to be accounting. However, since standard representationalist semantics does not aspire for an account of the concepts or properties expressed by expressions like "red," there is simply no need to specify the meanings of expressions like "red" in such a way that does not presuppose a grasp of these very meanings. Inferentialism, on the other hand, has more ambitious aspirations. Inferentialism is a *semantic theory* that aspires to be an *account of conceptual content*. It aspires to be what Dummett calls a *full-blooded* theory of meaning. The distinction, in short, is this:

> **Modesty:** A *modest* theory of meaning does not aim to give any non-circular account of the contents expressed by the basic expressions of the language.
>
> **Full-Bloodedness:** A *full-blooded* theory of meaning aims to give a non-circular account of the contents expressed by the basic expressions of the language.

In general, a representationalist model-theoretic semantics only aspires to be a *modest* theory of meaning, whereas an inferentialist theory aspires to be a *full-blooded* theory of meaning. Why should one aspire to a full-blooded theory of meaning at all? Let me briefly state what I take to be the two most important reasons.

One of the main reasons for developing a full-blooded theory of meaning has to do with what one takes the relationship between concept possession and language learning to be. A modest theory of meaning will not be of any help in explaining how it is that speakers come to possess the concepts expressed by the expressions of the language. Indeed, if one's theory of language includes only a representationalist theory of meaning, the complementary theory of language learning will *presuppose* that speakers grasp the contents expressed by the expressions of the language for which one has a representationalist semantic theory. For instance, coming to know the meaning of "red," on the representationalist picture sketched above, will presumably involve coming to match up the expression "red" with the property it represents, namely, the property of being red, which is formally represented in the semantic theory by the function given above. But how should we understand the grasp of these properties that speakers of the language must have in order to learn the meanings of the expressions of the language? In response to this question, a representationalist semantic theorist might just shrug their shoulders and say "Not sure, but that's not a question of a semanticist! That's for philosophers of mind or epistemologists." Full-blooded theorists of meaning, by contrast, take it that this question in philosophy of mind and epistemology is to be addressed *through* the development of a full-blooded semantic theory. A full-blooded theory of meaning, in contrast to a modest theory, promises to *give an account* of speakers' grasp of the conceptual contents expressed by the expressions of the language in terms of their coming to know the master the rules governing the use of those expressions. Thus a full-blooded account of meaning is, at once, an account of concept possession. On inferentialist theory in particular, one can say that it is through mastering the specifically inferential rules governing the use of expressions that one comes to acquire the a grasp on the conceptual concepts expressed by those expressions.

The above motivation is, in effect, an epistemological one, concerning the account of our grasp of conceptual content. A further, closely related but distinct motivation for full-bloodedness is more metaphysical one, concerning what these contents themselves are. The contents expressed by sentences and predicates are traditionally metaphysically contentious entities. In particular, sentences express *propositions* and predicates express *properties* or *relations*. One essential feature of such conceptual contents is that they are *general*. There are many particular ways for the world to be and be such that Mars is red. Likewise, there are many particular things, different in many particular ways, that are all such that they are red. Such entities are *appealed to* in the context of a representationalist semantic theory. How should we understand what they are? Once again, in response to this question, a representationalist semantic theorist might just shrug and say "Not sure, but that's not a question for a semanticist! That's a question for metaphysicians." However, once again, a full-blooded theorist of meaning can maintain that that this question of metaphysics is to be addressed *through* the development of a full-blooded semantic theory. Indeed, perhaps Wilfrid Sellars's main motivation in developing inferentialism was to *give an account* what properties are in terms of the rules governing the use of sentences. What it is for something to be red, say, is understood in terms of the norms that one binds oneself by in using the term "red." On the view developed by Sellars, one arrives at an account of such metaphysical entities as properties by transposing an inferentialist theory from *normative* vocabulary to *alethic modal* vocabulary. By the end of this book, we will have gathered the resources to provide such an account of properties by transposing the inferentialist account put forward here. In some sense, that's the biggest payoff of the theory.

Now, I have talked thus far about modest vs. full-blooded accounts of the meanings of so-called *content words* like "red" or "yellow." Typically, however, a representationalist semanticist won't take themselves to really be giving an account of the meanings of such expressions, so pointing out that they don't give a substantive account of the meanings of such expressions is not really an *objection* to the theories they promotes. Often, however, such semanticists still *do* think that they are giving an account of the meanings of so-called *function worlds* like "and" and "not," for instance, in terms of the set-theoretic operations of intersection and complementation. Even here, however, we can still plausibly argue that a representationalist semanticist doesn't really give substantive account of these words. As we showed in the previous section, a representationalist semanticist will typically specify the significance of expressions like "and" and "not" in terms of set-theoretic operations like intersection and complementation. Specifying the singificance of these logical words in this way, we can point out that, while the semantic clauses *themselves* don't explicitly use "and" and "not," the standard specifications of what these set-theoretic operations actually are *do* use the words "and" and "not":

Intersection: $X \cap Y = \{a \mid a \in X$ and $a \in Y$
Union: $X \cup Y = \{a \mid a \in X$ or $a \in Y$
Complementation: $X - Y = \{a \in X \mid$ it's not the case that $a \in Y\}$

Accordingly, we may plausibly be skeptical as to whether a representationalist semanticist really *has* provided a non-circular account of even logical expressions like "and" and "not." Now, once again, it is of course open to the representationalist semanticist to maintain that they are not really aiming to give an *substantive account* of the meanings of logical words like "and" and "not." But, once again, the full-blooded theorist will respond that they *are* giving an account of the meanings of these expressions, and they can appeal to their account in order to explain the grasp of conjunction and negation that speakers have in terms of their mastery of linguistic rules.

Historical Note:

The account of language learning that I suggested the representationalist might appeal to, where learning the meaning of a word like "red" is matching it up with a property of which one already has a grip, constitutes

one of the main targets of Wittgenstein's *Philosophical Investigations*, known as the "Augustinian Picture" of language, due to Wittgenstein's quoting a description from Augustine's *Confessions* as exemplary of this conception. One of the main positive upshots of the *Investigations* is an alternate conception of meaning of language where the primary notion is not *representation* but *use*. As Wittgenstein says, for a majority of our deployments of the term "meaning," we can maintain "the meaning of a word is its use in the language" [11, 25]. Now, though Wittgenstein himself never intended to turn this suggestion into a full-fledged *use-based theory* of linguistic meaning (Wittgenstein himself was not one for philosophical theories), it inspired generations of philosophers influenced by his work to try to do just that.

Influenced by Wittgenstein, Michael Dummett was crucial in advancing the notion of a "use-based theory of meaning" in the second have of the twentieth century, and spelling out an developing such a theory in detail. As we'll see, Dummett's name will come up again and again in the following pages, as he was foundational in formulating many now commonplace notions of "proof-theoretic semantics."

1.4 Logical Inferentialism

The most widely accepted and technically well-developed form of inferentialism is a restricted form of inferentialism known as *logical inferentialism*, originally suggested by Gerhard Gentzen [4] and notably developed by Daq Prawitz. As the name suggests, this is inferentialism applied to give an account of the meanings of the logical constants, expressions like "and," "or," "not," "all," and so on. According to logical inferentialism, the meanings of logical connectives like "not," "and," "or," and so on can understood in terms of the inferential rules governing their use, canonically specified as rules in a certain kind of formal proof system. For example, the meaning of **and** might be understood in terms of the fact that its use is governed by the following rules in a *natural deduction* system:

$$\frac{A \quad B}{A \textbf{ and } B}\ \wedge_I$$

$$\frac{A \textbf{ and } B}{A}\ \wedge_{E_1} \qquad\qquad \frac{A \textbf{ and } B}{B}\ \wedge_{E_2}$$

Natural deduction is likely the sort of proof system you've encountered in an introductory logic course. In such a course, you likely used natural deduction as a tool to show that various arguments, with certain premises and a certain conclusion, are *valid*. A proof system like natural deduction is indeed helpful as such a tool. In this book, however, our concern with natural deduction and other poof systems will not be to *prove things*, but, rather, as a way of clearly explicating the inferential rules that we take to be constitutive of the meanings of the logical connectives. In the rules shown above, we associate **and** with an *introduction rule* specifying the inferential *conditions* for asserting a sentence of the form A **and** B and a pair of *elimination rules*, specifying the inferential *consequences* of asserting a sentence of the form A **and** B. The basic thought of logical inferentialism is that mastery of these rules is sufficient for grasp of the concept of conjunction.

Natural deduction, originally introduced by Gentzen in his 1935 Ph.D. dissertation, is the type proof system most commonly appealed to by logical inferentialists in articulating the meanings of the logical connectives in terms of the inferential rules governing their use. However, it is not the only type of proof system Gentzen introduced there. He also introduced different type of proof system called the *sequent calculus*. In the sequent calculus, the objects that the rules tell you how to infer between are not *single sentences*, but, rather *sequents*, which say that some sentences follow from others. In this kind of proof system, the rules for **and** might be given as follows:

$$\frac{X, A, B \vdash C}{X, A \textbf{ and } B \vdash C}\ \mathrm{L}_\wedge \qquad\qquad \frac{X \vdash A \quad X \vdash B}{X \vdash A \textbf{ and } B}\ \mathrm{R}_\wedge$$

Here, rather than giving **and** introduction and elimination rules, we give it *left* and *right* rules. The left rule codifies its role as a *premise*. In particular, it tells us that if we can infer some sentence C from A along with B (along with some set of auxiliary premises X), then we can infer C from A **and** B. The right rule codifies the role of a sentence of the form A **and** B as a *conclusion*. In particular, it tells us that if we can conclude A from some set of premises X and we can also conclude B from X, then we can conclude A **and** B from X.

The next four chapters will concern logical inferentialism. We will focus, specifically, on the task of inferentially accounting for the meanings of the logically connectives, as these meanings are understood in the context of *classical* logic. The reason for this focus is, first, that classical logic is, by far, the most prominent logic of concern to logicians, and, second, accounting for the meanings of the connectives of classical logic in inferentialist terms poses certain well-known challenges that it will be productive to overcome. In the next chapter, we will introduce natural deduction as a proof system for accounting for the meanings of the logical connectives and present a famous problem having to do with negation. In Chapter 3, we will consider the sequent calculus as a proof system for accounting for the meanings of the logical connectives, and present a different problems having to do with the interpretability of the system, which will feature multiple conclusions. Ultimately, these problems will lead us to an approach to logical inferentialism known as *bilateralism*, where the opposite speech acts of *assertion* and *denial* will be treated as equally basic in giving an inferentialist account of the meanings of the logical connectives. In Chapters 4 and 5, we will introduce bilateral systems (both natural deduction and sequent calculi) in which sentences are positively or negatively *signed* to express their assertion or denial. Thus, rather than just having sentences like A or B, we'll have signed sentences like $+\langle A\rangle$ or $-\langle B\rangle$, expressing the assertion of A or the denial of B. The bilateral framework we end up with by the end of Chapter 5 will then serve as our basis going forward to expand inferentialism from just a theory about the meanings of *logical* expressions to a theory about the meanings of *all* expressions, logical and non-logical alike.

Historical Note:

Logical inferentialism is typically traced back to the work of Gerhard Gentzen, arguably the greatest logician

of the 20th century after only Gödel (whose legacy is hard to top). Gentzen more or less singlehandedly developed the field of proof-theory, and he used proof-theoretic methods the developed to prove a number of important mathematical theorems, perhaps most notably a consistency proof of Peano Arithmetic using transfinite induction. Beyond these mathematical achievements, however, his development of proof-theory laid the foundation for logical inferentialism. Gentzen influentially stated, in the context of natural deduction, that "The introductions represent, as it were, the 'definitions' of the symbols concerned" (80). Since Gentzen, this idea that the rules of a proof system constitute "definitions" of the connectives they govern has been taken seriously and developed into the logical inferentialist theories that are put forward today.

1.5 Global Inferentialism

Global Inferentialism is the ambitious form of inferentialism endorsed most notably by Wilfrid Sellars and Robert Brandom. According to global inferentialism, not just logical expressions like "and" but also non-logical expressions like "red," "bird," and "gravity" can be articulated in terms of the rules governing their use in inferences. The basic idea proposed by Sellars is that by expanding the notion of inference rules beyond *logical* inferences to include *material* inferences as well, we can inferentially account not just for the meaning of logical "function words" but also for the meanings of these "content words." The basic distinction between these two kinds of inference might be put as follows:

> **Logical Inference:** Holds in virtue of the *logical form* of the sentences involved, and so can be seen to be good in abstraction from the particular content of the sentences at issue.
>
> **Material Inference:** Holds in virtue of the *material content* of the sentences involved, and so cannot be seen to be good just by looking at the form of the sentences at issue.

So, whereas the inference from "x is red and x is round" to "x is red" is a logical inference, the inference from "x is red" to "x is colored" is a material inference. If we represent the logical forms of these sentences, in abstraction from the content, the first is an inference of the form $A \wedge B \vdash A$, which can be seen to be good at this level of abstraction, whereas the second inference is simply of the form $A \vdash B$, which cannot be seen to be good without looking into what sentences A and B actually are.

Though logical inferentialism is still a contentious philosophical and logical thesis, it is not nearly as contentious as global inferentialism. Very few philosophers or logicians would really claim to be global inferentialists. Still, global inferentialism is a much more exciting philosophical thesis than logical inferentialism since it promises to be an account of conceptual content *in general*. On this account, one can maintain that *what it is* for a word to express a certain concept just is for that word to be governed by a particular set of inferential rules. Moreover, one can maintain that acquiring concepts in general, be they formal or material, *just is* coming to master the inferential rules governing their use. Finally, only global inferentialism provides the promise of providing an account of properties and relations in general, such as the property of being red or the relation of being larger than, in terms of the inferential rules governing the use of predicates like "red" or "larger than." So, if global inferentialism can be made to work, there is quite a big philosophical payoff. However, there are serious challenges facing anyone wishing to develop a global inferentialist theory of meaning.

To briefly consider some of the challenges facing a global inferentialist, consider an expression like "red." Once again, on the representationalist, picture, the meaning of this expression is simply modeled as a function that picks out the red things in every possible world. Such a function says nothing about *what it is* for something to be red. An inferentialist specification of the meaning of "red" will at once be an answer the question of what it is for something to *be* red. That is, in giving an account of the meaning of "red," we give an account of the concept of being red. What, then is it for something to be red? If genuinely posed with this question, a very promising way to start answering it by saying what something's being red *entails* and *excludes*. Saying such things is one clear way of articulating what something's being red *amounts to*, what the *significance* of something's being red is. For instance, being red is a way for something to be *colored*, so if something's red, it can't possibly be gray or black. Moreover, if something is red and something else is pink, then the first thing must be darker than the second thing. Such statements articulate what it is for something to be red *holistically*, in terms of the relations of necessitation and exclusion that this content bears to other contents. An formally inferentialist theory will provide, schematically, a precise way of specifying such inferential relations so as to account for the content of "red" in inferential terms. Using the bilateral notation mentioned above, we might specify these inferential relations as follows:

$$+\mathbf{red}(x) \vdash +\mathbf{colored}(x)$$

$$+\mathbf{red}(x) \vdash -\mathbf{gray}(x)$$

$$+\mathbf{red}(x), +\mathbf{pink}(y) \vdash +\mathbf{darker}(x, y)$$

The inferentialist thought is that material inferential relations like these articulate the content of "red."

Of course, it's not the job of the inferentialist logician to actually spell out, in complete detail, the total set of inferential rules for any particular set of expressions. That would take a very long time, and it would be very boring. Rather, the inferentialist logician's job is to consider the different challenges in accounting for different kinds of inferential relations that play a role in constituting content and develop a system capable of accommodating them so that the total set of inferential rules constitutive of the content of any particular expression *could* in principle be spelled out in the context of that system, even though no one is actually going to do that. To get a sense of the sort of challenges the inferentialist logician will encounter, note the first that, if we really want to account for the content of "red," we cannot simply stick with color inferences. In order to account for the actual significance of "red" as opposed, for instance, to "green," we need to have some inferences pertaining to the actual things that are red. Thus, for instance, we might take into consideration facts such as that stop signs are red, that blood is red, and that roses are red. Consider just the last fact, the expression of which constitutes the opening line of many childhood poems. In general, from something's being a rose, it follows that it is red. So we have the following inferential rule:

$$+\textbf{rose}(x) \vdash +\textbf{red}(x)$$

Notably, however, unlike the inferential rules specified above, articulating the content of "red" in terms of its relation to other color words, this inferential rule is a *defeasible* rule. If, for instance, you add to the premise set the claim that x symbolizes friendship, it is no longer a good inference.[2] That is, though we have the above rule, we also have the following non-rule:

$$+\textbf{rose}(x), +\textbf{symbolizes friendship}(x) \nvdash +\textbf{red}(x)$$

This fact presents the inferentialist logician with a challenge, for it means that they must reject the structural rule known as *Monotonicity*, which says that adding premises to an inference never defeats the goodness of an inference. While this rule clearly holds for logical inferences, and so logicians are generally happy to assume it in developing logical systems, for many (indeed, perhaps most) material inferences, it does not hold. Thus, in spelling out global inferentialism, we will need to develop a formal system that is distinct from the sorts of systems widely used by logicians which impose this structural rule.

A further potential challenge has to do with accounting for the fact that some words are essentially such that they can be used *non-inferentially* in perception. For instance, it seems essential to the meaning of "red" that one can apply the expression "red" to some object upon *seeing* that it is red. How does one capture *that* fact in inferential terms? The vast majority of inferentialists have, in effect, given up inferentialism in response to this question, maintaining that, in addition to material inferences between uses sentences and other uses of sentences, an "inferentialist" theory must also contain "quasi-inferences" obtaining between perceptual circumstances and uses of language. The approach taken in this book, by contrast, follows Sellars and Brandom in aiming to account for such facts in genuinely inferential terms, in terms of special sorts of material inferences. In particular, we can inferentially account for the fact that red things can be seen to be red with a material inference along the following lines:

[2]Note that, whereas red roses are symbols of roses, yellow roses are symbols of friendship.

$$+\textbf{red}(x), +\textbf{adult human}(y), +\textbf{looking at}(y, x), \vdash +\textbf{sees that}(y, \langle x \text{ is red}\rangle)$$

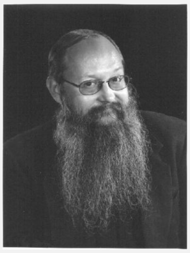
Note first that this is a *defeasible* inference, like the one from **bird** to **flies**. For instance, if you add in the premise **color blind**(y), the inference is no longer good. That, of course, is no problem since our system will already be designed to accommodate such inference. More notably, this inference involves the ascription of a *propositional attitude verb*, "sees that," which takes sentential objects, and will be inferentially related to other propositional attitude verbs such as "believes that," "knows that," and so on. Accommodating this inference, then, will require a system for attributing propositional attitudes, and the various further challenges that come with that.

I have given you these examples here just to give you a basic sense of the shape our global inferentialist theory will take. By the end of this book, you will have the resources to actually spell out, in formal detail, not just an account of the meanings of logical words like "and" and "or" in terms of the rules governing their use, but also the basic resources needed to spell out an inferentialist account of the meaning of content words like "red," "bird," and "gravity" in terms of the inferential rules governing the use of sentences containing them. Of course, the formal development of global inferentialism is still in its very early stages, and so what is provided in this book will be far from the full story. Still, the hope is that this book will give you enough of a foundation so that you can go on to further develop inferentialist approach to meaning yourself, whether that is building on the framework put forward here or developing your own inferentialist framework that is capable of going far beyond what the framework put forward here is capable of.

Historical Note:

As noted above, the sort of inferentialist theory of meaning pursued in this book is owed primarily to Wilfrid Sellars and Robert Brandom. Sellars was hugely influenced by Carnap, but rather than focusing on Carnap's development of possible worlds semantics (though he did discuss possible worlds in some early

writings), Sellars was more taken by Carnap's attempt to systematically articulate the rules. Sellars saw his early work as an expansion of Carnap's program, moving it beyond strictly *logical* rules and applying it to *material* rules as well to give a complete account of the contents expressed by the expressions of a language in terms of the inferential rules governing their use. One of the most important applications for Sellars, like Carnap, is the analysis of talk of *abstract entities* like properties and relations that an inferentialist account is able to provide. On Sellars's account, talk of abstract entities is a way of covertly expressing linguistic rules.



Hugely influenced by Sellars, Brandom sought to provide serious philosophical grounding to an inferentialist theory of content. His 1994 book *Making It Explicit* is the first book-length development of global inferentialism in the analytic tradition, and it's a *long* book-length development at that. However, though *Making It Explicit* is very philosophically rich, Brandom never actually provides anything like a formal framework in which one can actually *do* inferentialist semantics. In the past few decades, Brandom and a number of collaborators (including myself) have been working to do this. Much of the material in this book is a result of collaborating with Brandom and others in our research group called "Research on Logical Expressivism," or ROLE (pun intended).

1.6 Exercises

1. In stating the representationalist model-theoretic approach I did not formulate a semantic clause for a conditional sentence of the form $A \rightarrow B$. Suppose, we interpret $A \rightarrow B$ as logically equivalent to $\neg(A \wedge \neg B)$ (or, equivalently $\neg A \vee B$). State a semantic clause to give it this meaning. Then draw a Venn Diagram showing its truth conditions.

2. Determine whether or not the following is a formal or material inference:

   a) Pittsburgh is east of Chicago. So, Chicago is west of Pittsburgh.

   b) If it's raining, the streets are wet. It's raining. So, the streets are wet.

   c) Alice's car is gray. Bob's car is black. So Bob's car is darker than Alice's car.

   d) Everything red is colored. The ball is red. So the ball is colored.