CHAPTER 9

Truth and Meaning

Robert Cummins

Introduction

DONALD DAVIDSON'S "Meaning and Truth," revolutionized our conception of how truth and meaning are related (Davidson 1967). In that famous article, Davidson put forward the bold conjecture that meanings are satisfaction conditions, and that a Tarskian theory of truth for a language is a theory of meaning for that language. In "Meaning and Truth," Davidson proposed only that a Tarskian truth theory is a theory of meaning. But in "Theories of Meaning and Learnable Languages," he argued that the finite base of a Tarskian theory, together with the now familiar combinatorics, would explain how a language with unbounded expressive capacity could be learned with finite means (Davidson 1965). This certainly seems to imply that learning a language is, in part at least, learning a Tarskian truth theory for it, or, at least, learning what is specified by such a theory. Davisdon was cagey about committing to the view that meanings actually *are* satisfaction conditions, but subsequent followers had no such scruples.

We can sum this up in a trio of claims:

Davidson's Conjecture

(1) A theory of meaning for L is a truth-conditional semantics for L.
(2) To know the meaning of an expression in L is to know a satisfaction condition for that expression.
(3) Meanings are satisfaction conditions.

For the most part, it will not matter in what follows which of these claims is at stake. I will simply take the three to be different ways of formulating what I will call Davidson's Conjecture (or sometimes just The Conjecture).

Davidson's Conjecture was a very bold conjecture. I think we are now in a

position to see that it is probably false, but I do not expect many to agree with me about this. Since the publication of "Meaning and Truth," truth-conditional semantics has been pretty much all the semantics there is. In the current climate, therefore, it is something of a challenge to get philosophers of language to realize that the Conjecture is not obviously *true*. Generations of philosophers have been trained to regard The Conjecture as a truism. What else could semantics be? Surely, to understand an expression, one must know the conditions under which it is satisfied!

Prior to Davidson, semantics, at least in Philosophy, was speech act theory: Austin, Grice, and their followers (Austin 1962; Grice 1957). That tradition either died, or was co-opted. Here is how the co-option went. The Grician program, in the hands of Bennett (1973, 1976), Bach and Harnish (1979), Lewis (1969), Shiffer (1981, 1982), and their followers, reduces linguistic meaning to intentional psychology—i.e., to propositional attitudes. Fodor (1975) , Schiffer (1981), and others then introduced what I call the representational theory of intentionality (RTI hereafter): the idea that an intentional attitude is a mental representation in a cognitive role—e.g., a belief is realized as a sentence in mentalese available as a premise in inference but not as a goal specification. So, meaning for public language reduces to the attitudes, and the attitudes reduce to cognitive psychology and a theory of mental representation. A theory of mental representation, in this tradition, is, in Fodor's words, supposed to tell us where truth conditions come from (Fodor 1987, 1990). And that brings us back to Davidson's Conjecture. Meanings for Mentalese are to be given by a truth-conditional semantics, and the content of a propositional attitude is just the truth-condition of its associated mental representation. Meanings for a natural language, then, are specified finally in terms of the truth conditions of the Mentalese constituents of the attitudes involved in linguistic communication.[1] Thus Grician speech act theory ultimately rests on truth-conditional semantics. The substantive content of Speech Act Theory was relegated to "pragmatics"—the business of distinguishing promises from threats, and specifying the contextual factors involved in determining truth-conditions.

Of course, you do not need to be a Grician about meaning to get to this point. All you really need is the view that understanding[2] an expression E in a language L requires a mental state—either a representation or an intentional attitude—that has the same content as E. This reduces the theory of meaning and understanding the expressions of a language—the semantics anyway—to the theory of mental content. You then assume that a theory of mental content assigns truth/satisfaction conditions to mental states, either directly, or via the RTI. And that brings you back to Davidson's Conjecture.

So the philosophy of language turned into truth-conditional semantics, and the philosophy of mind labored to explain how mental representations could come to have the satisfaction conditions required. Thus it is that "Meaning and Truth" set the agenda for the philosophy of language and the philosophy of mind, linking the two tightly together in the process.

The link is more important that it might first appear. Once you have a Davidsonian story about the semantics of natural language, it is nearly irresistible to conclude that intentional states or mental representations (or both) must have a truth-conditional semantics as well. How else could we hope to get a grip on how it is possible to mean and understand the expressions of a language? If the meanings of linguistic expressions are satisfaction conditions, and someone knows the meanings of those expressions, then surely they know satisfaction conditions for those expressions. The knowledge is tacit, of course, but can be tapped by suitable queries about what is "intuitively" true under some specified set of hypothetical or actual circumstances. This is how we get the conclusion that mental representation must be, "classical" (Fodor and Pylyshyn 1988; Fodor and McLaughlin 1990). It is worth setting this out explicitly.

- Davidson's Conjecture: the meaning of a linguistic expression is a satisfaction condition for it.
- To understand a linguistic expression that means M, you must be able to token a mental representation that means M. (For example, to have the thought that p you must be able to token a mental representation that means that p.)
- Hence, mental representations must have a truth- conditional semantics, i.e., they must be "classical."

This inference from the Conjecture to the nature of mental content carries a price.[3] To see what it is, we need to begin with a basic constraint on any theory of linguistic meaning.

Communicative Constraint: The meaning of a natural language expression is whatever it is you have to know to understand that expression.

What I have just called the communicative constraint on linguistic meaning says, in effect, that linguistic meanings are whatever it is that have to be grasped or possessed for linguistic communication to be successful. Ultimately, a theory of meaning for natural language must dovetail with the psychology of linguistic communication.[4]

We can now see why the inference from Davidson's Conjecture to the na-

ture of mental representation could be pricey. There are good reasons to think that the mental structures required for language understanding do not have a truth-conditional semantics. It is going to be the burden of this chapter to argue this point. If you accept the point, and you accept the Communicative Constraint on linguistic meaning, you will think that a theory of language understanding will make no use of truth-conditional semantics. It doesn't follow from this that natural languages don't *have* a truth-conditional semantics. But it does follow that there is no good reason to think that a truth-conditional semantics for natural language will have any place in a mature psycholinguistics.

So here is the bottom line: I think that Davidson's Conjecture is a mistake. I think that truth has little to do with meaning. Or rather, so that we won't simply slide into arguing over the word, I think that truth has little to do with speaking and understanding a language.

Communicative vs. Referential Meaning

Let's begin with some terminology. By the *communicative meaning* of a term in a language I mean whatever you have to have in your head to understand it.[5] By the *truth-conditional meaning* of a term in a language I mean its satisfaction condition, or its role in generating one in the pragmatic context of some particular production of it. We can now express the central question thus:

- Are communicative meanings truth-conditional meanings?

OK. So what do you have to have in your head to understand, say, 'elevator'? Well, you have to have a more or less adequate concept of an elevator. But this just names the problem. What do you have to have in your head to have a concept of elevators? I think it is pretty clear that what you need is some basic knowledge of elevators. If you ask someone what 'elevator' means, they will tell you what an elevator *is*. They might, if they are very forthcoming and articulate, say something like this:

> Imagine a little room like a closet that moves up and down in a vertical shaft in a building. You get in on one floor, and the thing moves up or down to other floors where you can get off. Faster and easier than stairs. I think it is done with pulleys. Modern ones are controlled with buttons inside, and you can summon it with a button by the door leading to it on any floor.

And they draw a diagram:

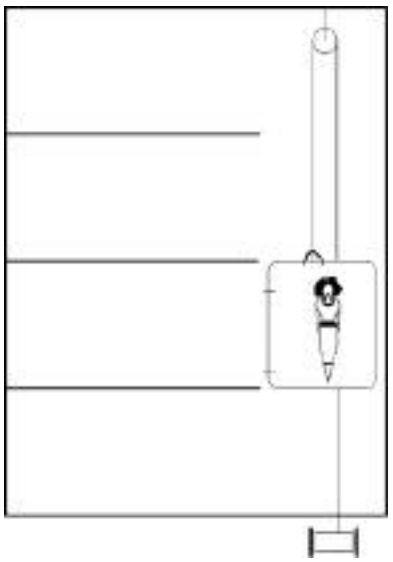

Figure 9.1. Drawing of an elevator

This much, I think, would be plenty in ordinary life or a psychology experiment to demonstrate that the "subject" has the (or a) concept of an elevator. And it would be enough precisely because it would demonstrate basic knowledge of elevators. So it seems clear that one can be said to have concepts in virtue of having a basic knowledge of their instances. If you know what an elevator is, you have the concept of an elevator. Moreover, if you ask someone what 'elevator' means, the same answer will do the trick. If, in answer to the question, "What does the word 'elevator' mean?" they demonstrate possession of a 'pretty good' concept of an elevator, then they know what 'elevator' means.

All of this knowledge one has that comes into play in connection with elevators is not just about elevators, of course. It is also about buildings and pulleys, for example. But the *topic* of the knowledge that one accesses when, as we say, one applies the concept of an elevator, is: *elevators.* Similarly, one can have an entire book about elevators. That book will also be about lots of other things, but the topic is elevators. I have no general analysis of what it is that determines the topic of a book or a body of knowledge. I don't think it is a very tight notion. Psychologically, the knowledge that gets accessed when a certain concept comes into play will vary from occasion to occasion and from person to person. My knowledge differs from yours, and my own is constantly changing. Moreover, which parts or aspects of my knowledge of a particular topic I happen to access on a given occasion will depend on the cues and on prior activation. But, given a reasonable amount of shared knowledge and stability over time, we can expect, in ordinary cases, a large overlap of core knowledge across persons and (reasonable) times.

On this view of things, the concept of a horse, and hence the communicative meaning of the word 'horse', is not a mental representation the reference of which is horses or the property of being a horse. It is, rather, a body of knowledge loosely identified by its topic. Just as a book about horses has horses as its topic, but not its referent, so a concept of horses has horses or the property of being a horse as its topic rather than its referent. With some trepidation, I'm going to sum this up by saying that a concept (of horses, say) is a *theory* (of horses),

the idea being that theories are organized bodies of knowledge that we identify in much the way we identify concepts—viz., by specifying a topic. One can have a theory of motion or a concept of motion; one can have a theory of pain or a concept of pain; one can have a theory of success or a concept of success. Theories, like concepts, are identified by their topics, not by their referents.[6] And they are, at least on the hoof, blessed with fuzzy boundaries that overlap other theories identified by other topics. Indeed, the identification of theories by topic, while useful, is a kind of heuristic in just the way I think the standard identification of concepts is a heuristic: it points you in the right direction if you want to look it up, but not much more.[7]

Concepts, conceived as (perhaps tacit) theories, are pretty clearly what you need to have in your head to understand terms such as 'elevator' and 'brown' and 'horse' and 'galloping'. They are also just what you need, along with the relevant sensory apparatus, to recognize elevators, horses, brown, and the gallop. And they are what you need to reason about such things. All of this is as it should be, since, when someone tells you not to ride the brown horse as he is likely to gallop, you don't want to avoid riding elevators. An immediate consequence of this view of concepts, and hence of communicative meanings, however, is the following:

> Concepts do not semantically combine in the way required by truth-conditional semantics.

The standard Tarskian combinatorics (Tarski 1956) suggests a mechanical process for combining a mentalese term for being striped with a mentalese term for being a fish, a process that yields a complex mentalese term for being a striped fish. But no Tarskian process will semantically combine a theory of stripes with a theory of fish to yield a theory of striped fish. Even more obviously, the denial of a theory of fish is not a representation applying to all nonfish in the way that the denial of a Mentalese term for fish is (or would be if there were such a thing), precisely, a term applying to the nonfish. Tarskian combinatorics are hopeless in connection with the sorts of psychological structures concepts must be to do their jobs.

This is an instance of a widespread problem. The kinds of mental representations that are subject to Tarskian semantics are what Fodor and Pylyshyn (1988) call 'classical' representations: language-like concatenations of arbitrary primitive symbols whose syntactic rules of formation are directly exploitable by truth-conditional combinatorics. No one would dream of trying to exploit Tarskian truth-theory to cope with the semantic complexity and productivity of pictures, maps, graphs, or activation vectors. It only works for language-like schemes. Yet there is little reason to think that classical, language-like schemes have any real representational virtues. This is because there are basically just three ways that arbitrary mental symbols of the Language of Thought variety

can enter into cognitive explanations: As *triggers* for procedures, as *cues* for stored knowledge, and as *constituents* of complex representations.

The point can be brought out by a simple example. You are asked to go milk the cow. You make a plan to carry out this request. Among your early subgoals is the subgoal to find the cow. You decide to look in the barn. When you get to the barn, you walk around inside looking for the cow. You look in a stall, and token a |cow|—a mental symbol that refers to cows. But just how did this cow recognition work? To recognize cows, you need to know something about them. You need, at least, to know how they look. A mental symbol does not contain any information about how cows look, and so it is not what psychologists would call a concept. You need to deploy your knowledge of cows in order to recognize a cow. It is your knowledge of cows, including tacit knowledge about the sort of retinal projections they tend to produce, that makes it possible for you to token a |cow| when you encounter a cow. So the Mentalese |cow| did no work for the object recognition system, its just signaled its output.

But that is not all. Having tokened a |cow|, where do you stand in the great scheme of things? The |cow| tokening *triggers* the next step in the plan. Now that you have located the cow and are on the spot, you need to locate the udder. Here, something like a picture of a cow, an image, say, would be very helpful, whereas a mental word is totally useless unless it happens to function as a retrieval *cue* for some stored knowledge about cows. Faced with actually having to deal with a cow, the burden therefore shifts again from the symbol to your stored knowledge, because the symbol, being arbitrary, tells you nothing about cows. So it turns out that it is not because you have a Mentalese term for cows that you get the milking done, it is because you have a route—activated by a cue—to something else, some stored knowledge about cows. Mentalese |cow|s could play a role in stored knowledge about cows only as pointers to it, or as constituents of complex representations—|cows have udders between their back legs|, for example—that are, on the Mentalese story, implicated in the possession of stored knowledge about cows.

I do not think this should come as any real surprise to LOTers, for I think the view is widespread among them that it is really stored knowledge that does the explanatory work anyway. But it is worth emphasizing that there is a big difference between appealing to the fact that one has a primitive mental symbol referring to cows, and appealing to the fact that one has a lot of knowledge about cows. LOT commits one to the view that representations of cows don't tell you anything about cows.

Perhaps it is not so bad that LOT entails that the representations that are satisfied by cows have only an indirect role in the explanation of cow cognition, for there are always mental sentences to tell us about cows. But let us just be clear about what LOT is committed to here: The view we have arrived at is that cognition is essentially the application of a linguistically expressed theory. All the

serious work gets done by sets of sentences that are internal tacit theories (ITTs) about whatever objects of cognition there happen to be. As far as cognizing cows goes, your |cow|s really don't matter; it is your ITT of cows that does the work.

But, of course, ITTs are not subject to Tarskian combinatorics. Indeed, it is pretty obvious that no psychological structure can play the roles characteristic of both a Tarskian term and concept. Concepts, for example, subserve object recognition. A concept of a fish (a FISH) is what enables one to recognize fish. To recognize fish, you need to know something about fish—you need a theory of fish, in short. Having a Mentalese term is of no use at all; you have to learn to token that term in response to fish, and that is just what knowing something about fish allows you to do, and what you cannot hope to do if you don't know anything about fish. Similarly, to understand the word 'fish', you need to know something about fish. Having a mental term, by itself, would be no help at all, since having a mental term referring to something is not the same thing as knowing anything about it. You cannot understand 'fish' if you do not have a FISH, and your understanding of 'fish' is exactly as good as your FISH.

Mental terms in a language of thought, if there is such a thing, have satisfaction conditions: something counts as a |fish| just in case it is satisfied by fish. Consequently, mental terms in a LOT would be subject to semantic combination: you can combine a |striped| and a |fish| and get a |striped fish|. But having a |fish| at your disposal does not, by itself, endow you with any knowledge of fish, and hence does not enable you to recognize fish, or understand the word, or reason about fish. Expressions in a LOT might have the same truth-conditional meanings as the expressions of a natural language, but activating (tokening) a LOT expression that is truth-conditionally equivalent to an expression in a natural language could not possibly constitute *understanding* that natural language expression. To repeat, the story has to be that the Mentalese terms cue the corresponding theories.

Mental Merging

I have been urging that communicative meanings are rather like theories. Since theories are not candidates for relevant sort of Tarskian combinatorics, it follows that a Tarskian truth theory cannot be a theory of communicative meaning. As I pointed out earlier, this does not refute Davidson's Conjecture, but it strips Davidson's Conjecture of most of its relevance to Cognitive Science. Even if a natural language could be fitted with a truth-conditional semantics, that would not help explain how it is learned or understood. Since natural language is a biological adaptation whose function is enabling communication—a fact philosophers of language sometimes forget and almost always neglect—the interest in such a semantics would be largely or completely orthogonal to the problem of understanding how we understand a language.

But if concepts do not have a Tarskian semantics, how do we combine our

understanding of 'brown' and 'horse' to get an understanding of 'brown horse'? Theories do not simply merge, and the denial of a theory of horses is not a theory of nonhorses. Davidson's Conjecture, and its implications for language understanding, gave us a story to tell about how our understanding of complex expressions could be constructed from our understanding of their constituents. What shall we put in its place?

This problem would need facing even if you believed in a language of thought with a truth-conditional semantics. For suppose you have uttered, 'The man holding the brown shoe is my brother,' and my language understanding system has constructed a truth-condition for it. What it has, in effect, is a Mentalese translation of your sentence, containing Mentalese terms like |man|, |brown|, |shoe|, and |holding|. We can assume, for the sake of argument, that each of these activates the corresponding concepts, |man|s cuing MANs, |brown|s cuing BROWNs, and so on. But this is a far cry from having a conception of the state of affairs expressed by your sentence. How does one build up that conception from MANs, BROWNs, SHOEs, and so on, together with the truth-conditional combinatorics? Building a |brown shoe| form a |brown| and a |shoe| does not automatically give you a BROWN SHOE.

It is glaringly obvious, once the question is raised, that symbolically represented theories are not subject to Tarskian combinatorics. Truth-conditional combinatorics, therefore, allows you to explain how the truth-conditional meaning for a complex expression can be built up from the truth-conditional meanings of its components and its syntax, but it leaves untouched how the communicative meanings of complex expressions could be built up from the communicative meanings of their components. A truth-condition for a complex expression provides no clue as to how one might build up the conception of the situation that expression so readily conveys to the mind of a mature speaker. We are thus led to ask whether there is some other way of representing the relevant knowledge—some nonlinguistic way of representing the knowledge involved in BROWNs and SHOEs, for example—which does allow the kind of relatively straightforward concept-merging that real-time language understanding so obviously requires.

In connectionist networks, long-term knowledge is stored in the connection weights. Whatever such a system knows about shoes and brown resides somehow in the pattern of connectivity and the associated weights.[8] It is, in the present state of play, a mystery how we should "read" a pattern of connection weights. No one knows how to take a verbally expressed body of knowledge and express it as a pattern of connection weights. Indeed, if John Haugeland is right, and I think he is, this is impossible (Haugeland 1990). According to Haugeland, different genera of representational schemes allow for the expression of characteristically different contents. Pictures and sentences are intertranslatable only in the very roughest way. We should expect the same for sentences and patterns of

connection weights. However, this message of incomensurability between verbal and connectionist representation is a *hopeful* message in the present context, because we know that the problem facing us has no ready solution—perhaps no solution at all—in its verbal form: logically combining verbally expressed theories, to repeat, has no hope of giving us what we want. This, perhaps, is enough to justify a bit of wild speculation in spite of our ignorance of the semantics of weight matrices.

Think, then, of a weight matrix as an encoding (doubtless idiosyncratic) of a kind of know-how. It might be knowledge of how to retrieve an item from memory given a cue of some sort. This is what we have in the famous Jets and Sharks network of McClelland and Rumelhart (1988). Or it might be knowledge of how to pronounce English text, as in Sejnowski and Rosenberg's NetTalk. Know-how, it seems, is naturally captured in a weight matrix. Can we think of concepts as know-how? Certainly. To possess the concept of a shoe is, to a first approximation, to know how to recognize one, to know how they are worn, and, if one is a linguistic creature, to know how to describe one. Knowing how to describe a shoe is, of course, know-how like any other. In particular, we should not assume that knowing how to describe a shoe requires a sort of "declarative memory," where this is conceived as a stored Mentalese description. The stored-description account has many failings, not the least of which is that we do not always describe the same thing in the same way. We get a more realistic account if we imagine a network that generates descriptions as outputs, with the description generated depending on the details of the input and the current state of activation—*set*, as it used to be called in psychology. In a similar vein, having a conception of the color brown is being able to recognize it, being able to give instances of brown things, being able to relate brown to other colors (e.g., darker than yellow and lighter than black), and so on.

Can we assemble the connectionist know-how that goes with SHOE and the connectionist know-how that goes with BROWN into the know-how that goes with BROWN SHOE? Notice that this is not a question in semantics at all, but a question about the mechanics of network building. We need a design that exploits the presence of a BROWN network and a SHOE network and generates, on the fly, and temporarily, a structure that exhibits the kind of know-how characteristic of BROWN SHOE possession.

It must be confessed that we are nowhere near to understanding how this might be done. But we do, I think, have a pretty good beginning on how the problem should be posed.

We start with a brief consideration of representation in connectionist networks, beginning with simple three-layer feed forward cases. Following Paul Churchland (1998), consider a network that learns to discriminate hillbilly families in terms of facial resemblance. Figure 2 depicts a simplified version of such a network, with the activation space at the hidden layer contracted to allow

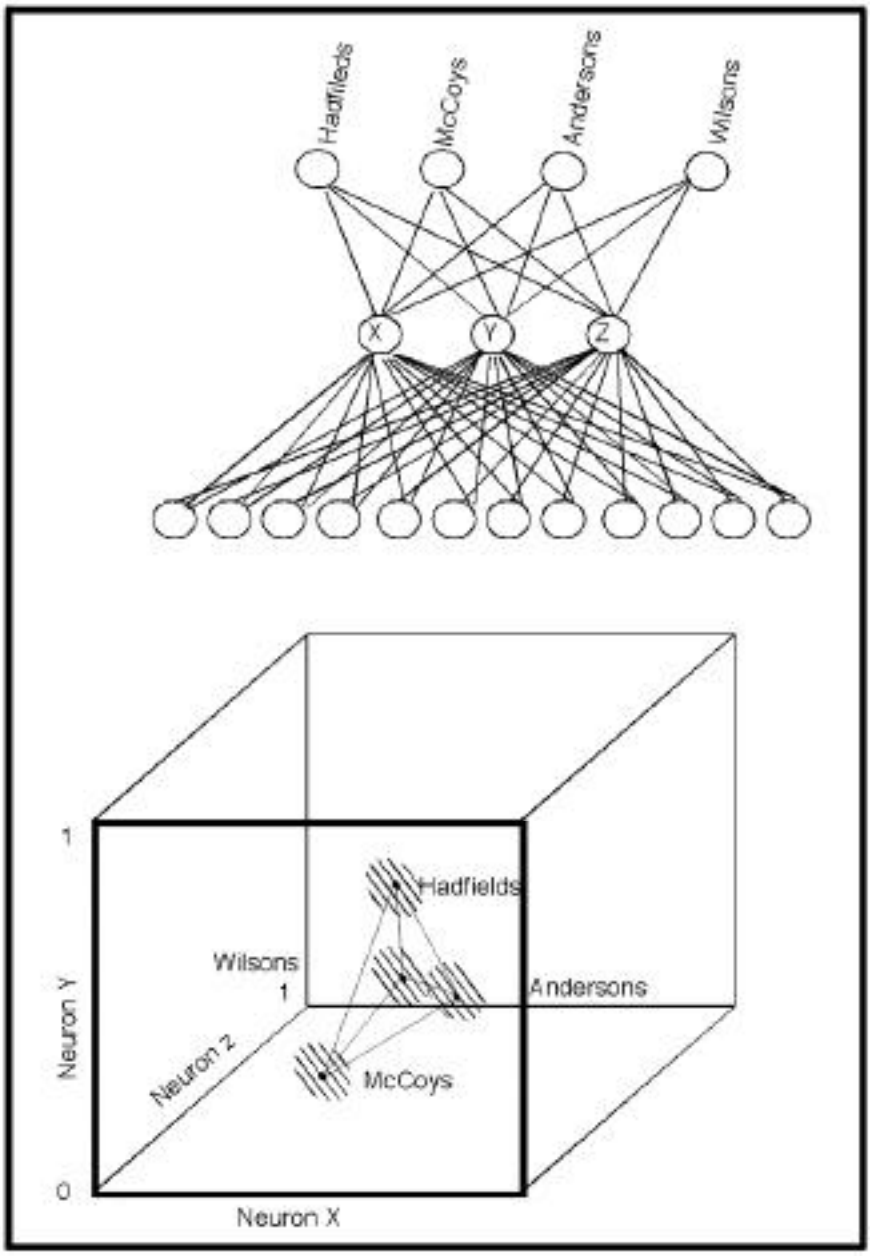

Figure 9.2. The network has acquired four prototype regions corresponding to facial family resemblance

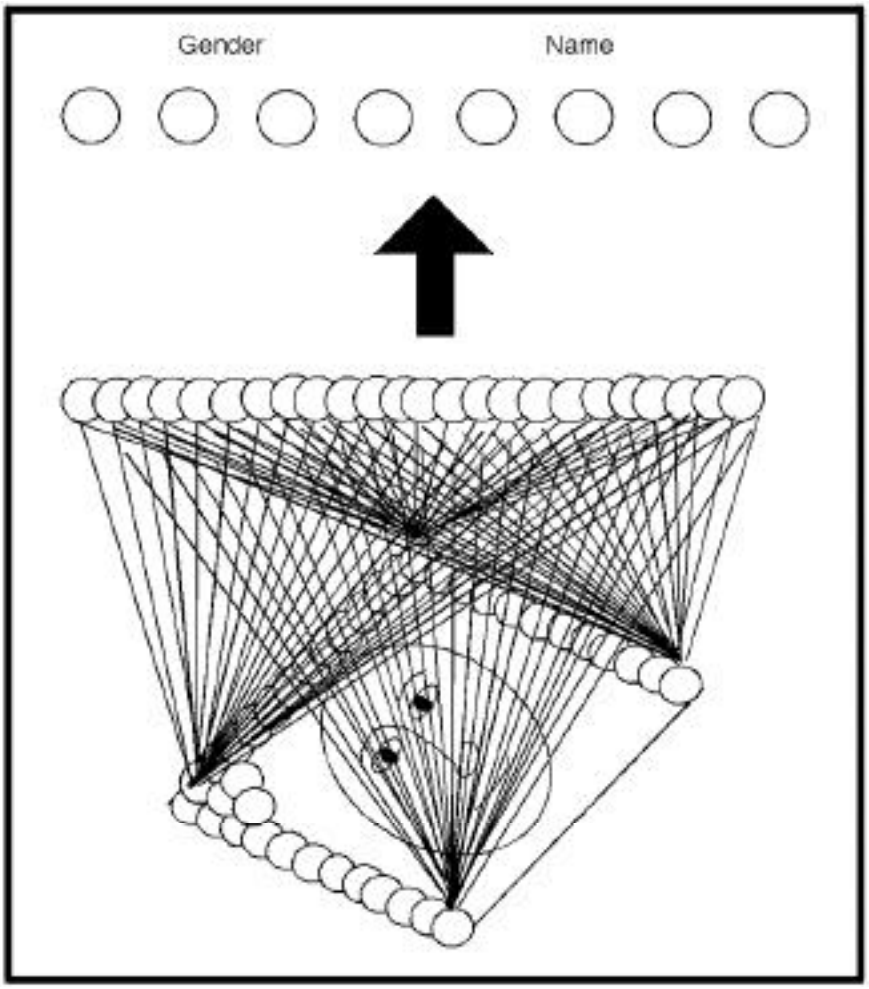

Figure 9.3. Schematic of Cottrell's face recognition network

three-dimensional illustration. The points in the space are what Churchland calls prototype points. They are centers of gravity around which cluster the points corresponding to related family members. They are a revealing way to represent the way that training the network partitions up the relevant activation space. The geometry thus revealed will be remarkably constant across different networks trained to the same task, including ones with differing input codings and even ones with differently dimensioned hidden layers (Laasko and Cottrell 2000). We are thus led to the idea that there is an objective structure to the relevant face space, and that trained networks discover this and represent it via an

isomorphic activation space. In such a space, it seems reasonable to think of the prototype points as something like individual concepts in a conceptual space. This perspective becomes more compelling as the networks face more complex tasks. Cottrell's tiny face recognition network (retina of 64 x 64 pixels, Figure 3) implicitly partitions its activation space in such a way that female faces tend to be closer to each other than to male faces and vice versa (Cottrell 1991a).

Simple recurrent networks of the sort pictured in Figure 4 pose a different case because they allow for dynamic representation. They are probably best conceived in terms of paths in activation space rather than points. This approach seems to work nicely for Elman's well-known grammar network, for example (Elman 1992).

Connectionist theory thus provides a compelling example of the kind of representation by structural similarity that I recommended in *Representations, Targets and Attitudes* (Cummins 1996). It provides representations that are structurally rich, representations that themselves guide cognition rather than func-

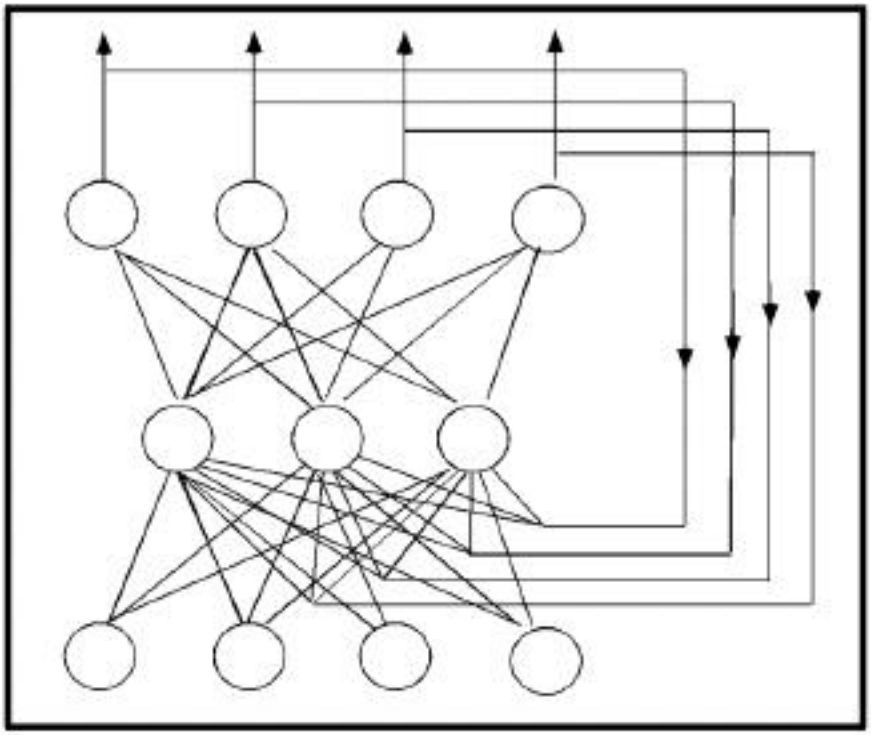

Figure 9.4. Simple recurrent network. Projections from the output layer to the hidden layer give the network a kind of short term memory of its immediate history

tion as mere indicators in detection. Unlike the representations posited by LOT theories, these representations are plausible candidates for concepts.

They are not, as yet, however, plausible candidates for the sort of fleeting merges that seem to underlie language understanding. No cross-network associations between, e.g., a color network and a shape network, will fill the bill here because, first, associations have to be learned, and, second, because they have to be unlearned to go away. A reference to yellow dogs early in your discourse makes it easier to understand talk of brown dogs later, not more difficult. There are powerful connectionist techniques for representing hierarchical bindings of the sort found in parse trees (Smolensky et al. 1992). It is tempting to suppose that vectors representing a parse could somehow be used to orchestrate the kind of conceptual liaisons we are after, but I think it is fair to say that no one currently knows how to do this.

The Communicative Function of Language

A novel conception of the function of language emerges from the foregoing discussion. Davidson's Conjecture implies that language is a medium for the expression of propositions and their constituents. It serves its communicative function when the hearer figures out what proposition the speaker expressed (or perhaps which the proposition speaker intended to express). The approach I have been urging implies that language is primarily in the communication business, and only secondarily, if at all, in the expression business. Sentences, on this view, are like recipes for assembling chunks of know-how into a know-howish conception of the speaker's communicative intention, and of the situation as the speaker conceives it. Sentences, in effect, tell you how to cook up a thought, where the thoughts thus cooked up are as different from words as are the cakes and pies from the recipes that tell you how to cook *them* up.

Viewed this way, it is possible—indeed, likely—that language can be used to communicate things it cannot begin to express, something poets and good novelists have always known. You can begin to get a sense of this by looking at the provision that language makes for "plug-ins." A plug-in, as every web browser knows, is an independent routine that your browser can "call" when needed, e.g., to decompress a downloaded file. Language uses demonstratives to construct docking places for these devices, as illustrated in Figure 5.

In your head, though, it is *all* plug-ins, a fact that has, I think, been obscured by the exaptation of language, especially written language, for expressive purposes quite foreign to its original biological function of facilitating communication in the service of social coordination. The expressive power of language is impressive, but hardly universal. It is, I think, much better at communicating thoughts than it is at expressing them. Failure to notice the distinction has led to the view that the only thoughts that can be communicated are the ones that can be expressed. When we put this together with Davidson's Conjecture, we

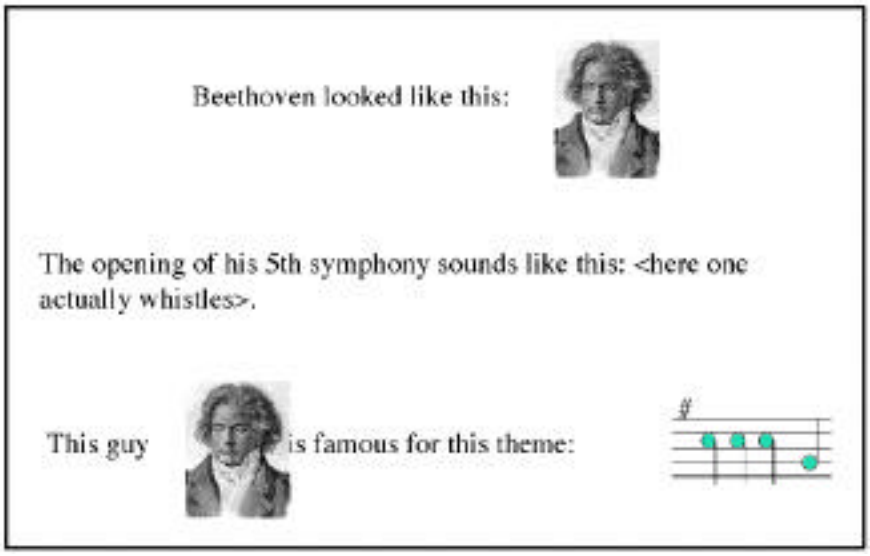

Figure 9.5. Linguistic expressions with plug-ins

get the result that the only thoughts that can be communicated are those that have truth-conditional contents—propositions, in short. It is a short step from this position to the widespread view that the only thoughts we can have are the propositional attitudes, and hence that there is no thought or mental representation whose content language cannot adequately express. In our hearts we all know this is wrong, but recent philosophical tradition has taught us to live with it or suffer professional extinction.

It is nearly universally assumed that the communicative meanings of linguistic utterances are the same as their representational meanings. The idea goes like this: I have the thought that p that I wish to communicate to you. I construct a sentence that means (representationally) that p, and you decode it—i.e., you figure out what its representational meaning is, and conclude that that is what I meant to tell you. This story could be right. But it is important that we not just assume it. To see that it isn't inevitable, imagine a communicative system that works like this: There are instructions that tell you how to assemble nonlinguistic representations—pictures, say—from elements—pixels—you have available. In this system, the instructions and the messages communicated need have no meanings in common. Language *might* work like that. Sentences might be recipes for assembling thoughts, or even images, in the heads of others. If so, then the truth-conditions of my sentences, if they have any, will tell us nothing about what I communicate. This is because I can communicate an accurate picture to you without saying anything true about the scene pictured. The truth-

conditions of my sentences yields the limelight to the accuracy of the thoughts or other cognitive states they help to assemble.

To get a clearer view of the sort of possibility I have in mind here, consider the following communication system. You have a set of numbered storage bins. In these are little model houses, trees, ponds, lawns, roadways, signs, street lights, etc. You also have a table with a grid marked on it, with rows numbered and columns lettered. You get instructions like this:

- Put an item from bin 23 on 7A
- Center an item from bin 14 on 8C
- Put an item from bin 12 on 8D–8H

The result will be a model village. You assemble this representation on the basis of instructions that are built from a vocabulary that is utterly incapable of expressing any of the things represented by the model. The signal system and the representations it helps to assemble are representationally disjoint.

This sort of example demonstrates the possibility of a communication system in which the meanings the communicative symbols communicate are not the meanings they have. Could this be true of natural language? We are, I think, already in a position to see that it very likely is true of natural language. The words 'house', 'tree', 'pond', and so on, do not express the knowledge that constitutes your understanding of houses, trees, and ponds. They are signals that activate that knowledge, bring it on line, and, somehow, orchestrate its assembly into a more or less unified conception.

Beyond the Propositional Attitudes

I used to think (Cummins 1996) that nonlinguistic schemes could express propositions. For example, I thought we could take pictures to express propositions by following Stalnaker (1984) in thinking of a proposition as a set of possible worlds. Since a picture will "hold" in some possible worlds and not others, it partitions the set of possible worlds, and hence expresses a proposition. I now think, however, that Haugeland (1990) was right: sentences and propositions were made for each other, and so we must look elsewhere for the contents of nonlinguistic propositions.

The striking thing about maps, diagrams, partitioned activations spaces, pictures, graphs, and other nonlinguistic representations is that they are not true or false, but more or less accurate. A sentence either hits its propositional target, or it fails. Nonpropositional representations, however, are better evaluated in terms of a graded notion of accuracy. Moreover, such representations are typically multidimensional. Pictures, for example, represent (relative) size, shape, color, and (relative) location simultaneously. The possibility thus arises that two pictures might be incomparable in overall accuracy, since one might do better

on some dimensions—size and shape, say—while the other does better on others—color and location.[9] The concepts of truth and falsehood, and the Tarskian combinatorial semantics we have come to associate with them, will be no help at all in understanding how these nonpropositional representations fit or fail to fit their targets. Representational meaning for nonpropositional representations will have to be understood in different terms, as will their semantic structures.

A consequence of the graded and multidimensional nature of many nonlinguistic representations is that they do not partition up the set of possible worlds in any neat way. What we get instead is a kind of shading along a number of interdependent dimensions. Since I cannot think of a more catholic notion of propositions than the one Stalnaker endorses, I have to conclude that most, perhaps all, nonlinguistic representations do not express propositions and are not true or false.[10] But they evidently do represent. They represent how their targets are, with greater or less accuracy, along various dimensions. If we really want to understand meaning, we need to understand not only the representation of propositions, but the graded and multidimensional representation of nonpropositional contents as well. And if we want to understand the kind of meaning that is involved in mental representation, and hence in language understanding, we had best understand the kind of representation effected by the sort of dynamic partitioning of neuronal activation spaces that our synapses learn to effect. It would amaze me if truth-conditional semantics had anything significant to offer to this crucial research problem.

NOTES

1. There is a missing step here: Grician stories provide only propositional contents, hence provide meanings for nothing smaller than a sentence. The Tarskian combinatorics, however, require satisfaction conditions for terms. See Cummins (1996) for a proposal about how to get the Tarskian combinatorics into a Grician picture.
2. I am going to use 'understanding' as short hand for 'meaning and understanding' or 'using and understanding.' The idea is to have a single word for whatever you need to be either party—speaker or hearer—in successful linguistic communication. Passive mastery and active mastery of language differ, with the former outrunning the latter, especially in young children, and this suggests that there is more to speaking the language than there is to understanding it. Still, you have to understand it to speak it, and it is at least plausible that whatever you have to add to understanding (passive mastery) to get active mastery, it isn't more *semantics.*
3. It ought to be darkly suspicious, too, since it is a license to do experimental cognitive psychology from the armchair. We begin by asking after the truth-conditions of propositional attitude sentences, and wind up with conclusions about the structure and contents of psychological states. For more on this theme, see Cummins 1991.
4. This need not be the case for artificial languages, I suppose, since these need not be primarily in the communication business. They may be primarily in the business of expressing truths, and rely for whatever communicative efficacy they have on their connections with natural languages.
5. For the picky: Of course, you need to be awake, and to be smarter than a post. What we want is what you have to add to the mind to enable understanding of some particular expression not previously understood.
6. I'm not sure what the referent of a theory would be. If you thought a theory was a set of

sentences, which I do not, then, perhaps, the referent of a theory would be a proposition, viz., the proposition expressed by a conjunction of sentences used to express the theory.

7. Psychologists, of course, have suggested a number of theories about the form our concepts take. (The classic review is Smith and Medin 1981. For a recent review of the literature, see Gelman 1996.) They all have in common, however, the idea that a concept of X is stored knowledge about X that mediates recognition of and reasoning about Xs. The dispute is over how that knowledge is stored and deployed, e.g., as a prototype or exemplar that is compared to instances in recognition and used to generate premises in inference, or as a frame, script, or semantic net. What you do *not* find in the psychological literature is the idea that concepts are terms in Mentalese that are satisfied by the instances of the concept in question. You do not find this because it wouldn't work, as we will see.
8. Mathematically, we could reduce this to weights alone, dealing with connectivity by setting the weights between disconnected nodes to zero. But it is more intuitive to think in terms of what is connected to what, and how those connections are weighted. This allows us to think of a number of more or less independent nets that are only sparsely connected to each other.
9. It seems likely that high accuracy on one dimension will often have to be paid for in lower accuracy in others, given limited resources. The eye, for example, gains considerable resolution and color information via foveation, but loses light sensitivity in the process. A map that shows all the streets of London on one page will be either too big to use in the car, or be viewable only with magnification.
10. Vagueness in language introduces problems that appear similar on the surface. Whether they are genuinely related to the kind of multidimensionality and gradedness we find in pictures, models, and graphs is not at all clear.

References

Alston, W. 1964. *Philosophy of Language*. Englewood Cliffs, N.J.: Prentice Hall.

Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.

Bach, K. and R. M. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Bennett, J. 1976. *Linguistic Behavior*. Cambridge: Cambridge University Press.

———. 1973. "The Meaning-Nominalist Strategy." *Foundations of Language* 10: 141–168.

Churchland, P. M. 1998. "Conceptual Similarity Across Sensory and Neural Diversity: The Fodor/LePore Challenge Answered." *Journal of Philosophy* 95: 5–32.

Cottrell, G. 1991a. "Extracting Fratures from Faces Using Compression Networks: Face, Identity, Emotions and Gender Recognition Using Holons." In D. Touretsky, J. Elman, T. Sejnowski, and G. Hinton (eds.), *Connectionist Models: Proceedings of the 1990 Summer School*. San Mateo, Calif.: Morgan Kaufmann.

Cottrell, G. and J. Metcalfe. 1991b. "EMPATH: Face, Emotion and Gender Recognition Using Holons." In R. Lippman, J. Moody, and D. Touretsky (eds.), *Advances in Neural Information Processing Systems*, vol. San Mateo, Calif.: Morgan Kaufmann.

Cummins, R. 1996. *Representations, Targets and Attitudes*. Cambridge, Mass.: MIT Press.

———. 1991. "Methodological Reflections on Belief." In R. Bogdan (ed.), *Mind and Common Sense*. Cambridge: Cambridge University Press.

———. 1979. "Intention, Meaning and Truth Conditions." *Philosophical Studies* 35: 345–60.

Davidson, D. 1967. "Truth and Meaning." *Synthese* 17: 304–323.

———. 1965. "Theories of Meaning and Learnable Languages." *Proceedings of the 1964 International Congress of Logic, Mathematics and Philosophy of Science*, 383–94. Amsterdam: North Holland.

Elman, J. 1992. "Grammatical Structure and Distributed Representations." In S. Davies (ed.), *Connectionism: Theory and Practice* (*Vancouver Studies in Cognitive Science*, volume 3). New York: Oxford University Press.

Fodor, J. 1990. "Psychosemantics, or Where do Truth Conditions Come From." In W. Lycan (ed.), *Mind and Cognition*. Oxford: Basil Blackwell.

———. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, Mass.: MIT Press.

———. 1975. *The Language of Thought*. New York: Crowell.

Fodor, J. and B. McLaughlin. 1990. "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work." *Cognition* 35: 183–205.

Fodor, J. A., and Z. W. Pylyshyn. 1988. "Connectionism and Cognitive Architecture." *Cognition* 28: 3–71.

Gelman, S. 1996. "Concepts and Theories." In R. Gelman and T. K. Au (eds.), *Perceptual and Cognitive Development. Handbook of Perception and Cognition*, second edition. San Diego: Academic Press.

Grice, P. 1957. "Meaning." *Philosophical Review* 66: 377–88.

Haugeland, J. 1990. "Representational Genera." In W. Ramsey, S. Stich, and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Englewood Cliffs, N.J.: Lawrence Erlbaum Associates.

Laasko, A., and G. Cottrel. 2000. "Qualia and Cluster Analysis: Assessing Representational Similarity Between Neural Systems." *Philosophical Psychology* 13: 46–76.

Lewis, D. 1969. *Convention*. Cambridge, Mass.: Harvard University Press.

McClelland, J., and D. Rumelhart. 1988. *Explorations in Parallel Distributed Processing : A Handbook of Models, Programs, and Exercises*. Cambridge, Mass.: MIT Press.

Sejnowski, T., and C. Rosenberg. 1987. "Parallel Networks that Learn to Pronounce English Text." *Complex Systems* 1: 145–68.

Schiffer, S. 1982. "Intention Based Semantics." *Notre Dame Journal of Formal Logic* 23: 119–59.

———. 1981. "Truth and the Theory of Content." In H. Parret and J. Bouraresse (eds.), *Meaning and Understanding*. Berlin: Walter de Gruyter.

Smith, E., and D. Medin. 1981. *Categories and Concepts.* Cambridge, Mass.: Harvard University Press.

Smolensky, P., G. LeGendre, and Y. Miyata. 1992. "Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition." Technology Report 92–08. Institute of Cognitive Science, University of Colorado.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, Mass.: MIT Press.

Tarski, A. 1956. "The Concept of Truth in Formalized Languages." In *Logic, Semantics and Metamathematics*. Oxford: Oxford University Press.

BLANK
194