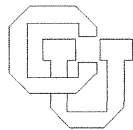


**The Constituent Structure of Connectionist Mental States:  
A Reply to Fodor and Pylyshyn**

**Paul Smolensky**

**CU-CS-394-88**



**University of Colorado at Boulder**

**DEPARTMENT OF COMPUTER SCIENCE**

**ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS  
EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR(S) AND DO  
NOT NECESSARILY REFLECT THE VIEWS OF THE AGENCIES NAMED IN THE  
ACKNOWLEDGMENTS SECTION.**



The constituent structure  
of connectionist mental states:  
A reply to Fodor and Pylyshyn

Paul Smolensky

CU-CS-394-88 March, 1988

Department of Computer Science &  
Institute of Cognitive Science  
University of Colorado  
Boulder, CO 80309-0430

Abstract

The primary purpose of this article is to reply to the central point of Fodor and Pylyshyn's (1988) critique of connectionism. The direct reply to their critique comprises Section 2 of this paper. I argue that Fodor and Pylyshyn are simply mistaken in their claim that connectionist mental states lack the necessary constituent structure, and that the basis of this mistake is a failure to appreciate the significance of distributed representations in connectionist models. Section 3 is a broader response to the bottom line of their critique, which is that connectionists should re-orient their work towards *implementation* of the classical symbolic cognitive architecture. I argue instead that connectionist research should develop *new formalizations* of the fundamental computational notions that have been given one particular formal shape in the traditional symbolic paradigm.

My response to Fodor and Pylyshyn's critique presumes a certain meta-theoretical context that is laid out in Section 1. In this first section I argue that any discussion of the choice of some framework for cognitive modeling (e.g. the connectionist framework) must admit that such a choice embodies a response to a fundamental cognitive paradox, and that this response shapes the entire scientific enterprise surrounding research within that framework. Fodor and Pylyshyn are implicitly advocating one class of response to the paradox over another, their critique is analyzed in this light.

Copyright © 1988 by Paul Smolensky.



# The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn

Paul Smolensky

*Department of Computer Science &  
Institute of Cognitive Science  
University of Colorado at Boulder  
Boulder, CO 80309-0430*

The primary purpose of this article is to reply to the central point of Fodor and Pylyshyn's (1988) critique of connectionism. The direct reply to their critique comprises Section 2 of this paper. In short, I argue that Fodor and Pylyshyn are simply mistaken in their claim that connectionist mental states lack the necessary constituent structure, and that the basis of this mistake is a failure to appreciate the significance of distributed representations in connectionist models. Section 3 is a broader response to the bottom line of their critique, which is that connectionists should re-orient their work towards *implementation* of the classical symbolic cognitive architecture. I argue instead that connectionist research should develop *new formalizations* of the fundamental computational notions that have been given one particular formal shape in the traditional symbolic paradigm.

My response to Fodor and Pylyshyn's critique presumes a certain meta-theoretical context that is laid out in Section 1. In this first section I argue that any discussion of the choice of some framework for cognitive modeling (e.g. the connectionist framework) must admit that such a choice embodies a response to a fundamental cognitive paradox, and that this response shapes the entire scientific enterprise surrounding research within that framework. Fodor and Pylyshyn are implicitly advocating one class of response to the paradox over another, and I wish to analyze their critique in this light.

## 1. The Paradox and several responses

In this section, I want to consider the question of what factors go into the decision about what cognitive modeling formalism to adopt, given the choice between the symbolic formalism and the connectionist formalism. I want to argue that the crucial move in deciding this question is to take a stance on the issue that I will refer to as "the Paradox of Cognition," or more simply, "the Paradox."

The Paradox is simple enough to identify. On the one hand, cognition is *hard*: characterized by the rules of logic, by the rules of language. On the other hand, cognition is *soft*: if you write down the rules, it seems that realizing those rules in automatic formal systems (which AI programs are) gives systems that are just not sufficiently fluid, not robust enough in performance, to constitute what we want to call true intelligence. That, quite simply, is the Paradox. In attempting to characterize the laws of cognition, we are pulled in two different directions: when we focus on the rules governing high-level cognitive competence, we are pulled towards structured, symbolic representations and processes; when we focus on the variance and complex detail of real intelligent performance, we are pulled towards statistical, numerical descriptions. The Paradox could be called, somewhat more precisely, The Structure/Statistics Dilemma.<sup>1</sup> The stance one adopts towards the Paradox strongly influences the role that can be played by symbolic and connectionist modeling formalisms. At least five noteworthy stances have been taken on the Paradox, and I will now quickly review them. I will consider each in its purest form; these extreme stances can be viewed as caricatures of the more subtle positions actually taken by cognitive scientists.

The first stance one should always consider when confronted with a paradox is *denial*. In fact, that's probably the most popular choice. The denial option comes in two forms. The first is to *deny the soft*. A more reputable name for this might be *rationalism*. In this response to the Paradox one insists that the essence of intelligence is logic and following rules—everything else is inessential. This can be identified as the motivation behind the notion of ideal competence in linguistics (Chomsky, 1965), where soft behavior and performance variability are regarded as mere noise. The fact that there is tremendous regularity in this noise is to be ignored—at least in the purest version of this stance.

The other denial stance is obviously to *deny the hard*. According to this view, rule following is really characteristic of *novice*, not expert, behavior; the essence of real intelligence is its *evasion* of rule-following (Dreyfus & Dreyfus, 1986). Indeed some of the strongest advocates of this position are connectionists who claim "there are no rules" in cognition.

If one rejects the denial options, one can go for the opposite extreme, which I'll call *the split brain*.<sup>2</sup> On this view, the head contains both a soft machine and hard machine, and they sit right next to each other. This response to the Paradox is embodied in talk about systems that have "connectionist modules" and "rule-based modules" and some sort of communication between them. There's the right, connectionist brain doing soft, squishy processing, and the left, von Neumann brain doing the hard rule-based processing. Rather than "the split brain," this scene of a house divided—right and left working side-by-side despite their profound differences—might better be called by its French name: *cohabitation*.

Advocates of this response presumably feel they are giving both sides of the Paradox equal weight. But does this response really grapple with the full force of the Paradox? In the split brain, there is a *hard line* that surrounds and isolates the softness, and there is no soft line that demarks the hardness. The softness is neatly tucked away in an overall architecture characterized by a hard distinction between hard and soft processing. The full force of the Paradox insists that the soft and hard aspects of cognition are so intimately intertwined that such a hard distinction is not viable. Not to mention the serious problem of getting the two kinds of systems to intimately cooperate when they speak such different languages.

The third approach to the Paradox is *the fuzzy approach* (Gupta, Ragade, & Yager, 1979). Here the basic idea is take a hard machine and coat its parts with softness. One takes a rule-based system for doing medical diagnosis and attaches a number to every rule that says how certain the inference is (Shortliffe, 1976; Zadeh, 1975, 1983); or one takes a set, and for every member in the set attaches a number which says how much of member of the set it is (Zadeh, 1965). In this response to the Paradox, softness is defined to be degrees of hardness. One takes the ontology of the problem that comes out of the hard approach, and one affixes numbers to all the elements of this ontology rather than reconceptualizing the ontology in a new way that intrinsically reflects the softness in the system.

On such ontological grounds, the fourth approach is starting to get rather more sophisticated. On this view, the cognitive machine is at bottom a hard machine; fundamentally, everything works on rules—but the machine is so complex that it *appears soft* when you look at it on a higher level. *Softness emerges from hardness*. This response to the Paradox is implicit in a comment such as

o.k., maybe my expert system is brittle, but that's because it's just a toy system with only 10,000 rules ... if I had the resources, I'd build the *real* system with  $10^{10}$  rules, and it would just be as intelligent as the human expert.

In other words, if there are enough hard rules sloshing around in the system, fluid behavior will be an emergent property.

In terms of levels of description, here's the picture. There's a level of description at which the cognitive system is hard: the lower level. And there's a level of description at which it's soft: the higher level. That's the sense in which this approach is getting more sophisticated: it uses *levels of analysis* to reconcile the hard and soft sides of the Paradox.

The question here is whether this approach will ever work. The effort to liberate systems built of large numbers of hard rules from the brittleness that is intrinsic to such rules has been underway for some time now. Whether the partial successes constitute a basis for optimism or pessimism is clearly a difficult judgment call.

The fifth and final approach I want to consider is the one that I have argued (Smolensky, 1988a) forms the basis of the proper treatment of connectionism. On this view, which I have called the *subsymbolic* approach, the cognitive system is fundamentally a soft machine that is so complex that it sometimes appears hard when viewed at higher levels. As in the previous approach, the Paradox is addressed through two levels of analysis—but now it's the lower level that is soft and the upper level that's hard: now *hardness emerges from softness*.

Having reviewed these five responses to the Paradox, we can now see why the decision of whether to adopt a symbolic computational formalism or a connectionist one is rooted in a stance on the Paradox. The issue is whether to assume a formalism that *gives for free* the characteristics of the hard side of the Paradox, or one that gives for free the characteristics of the soft side. If you decide not to go for combining both formalisms (*cohabitation*), but to take one as fundamental, then whichever way you go, you've got to either *ignore* the other side, or *build it* in the formalism you've chosen.

So what are the possible motivations for taking the soft side as the fundamental substrate on which to build the hard—whatever hard aspects of cognition need to be built? Here are some reasons for giving the soft side priority in that sense.

- A fundamentally soft approach is appealing if you view *perception*, rather than *logical inference*, as the underpinning of intelligence. In the subsymbolic approach, the fundamental basis of cognition is viewed as categorization and other perceptual processes of that sort.
- In overall cognitive performance, hardness seems more the exception than the rule. That cuts both ways, of course. The denial option is always open to say it is only the 3% that isn't soft that really characterizes intelligence, and that's what we should worry about.
- An evolutionary argument says that the hard side of the cognitive paradox evolved later, on top of the soft side, and that your theoretical ontogeny should recapitulate phylogeny.
- Compared to the symbolic rule-based approaches, it's much easier to see how the kind of soft systems that connectionist models represent could be implemented in the nervous system.
- If you're going to base your whole solution to the Paradox on the emergence of one kind of computation from the other, then it becomes crucially important to be able to analyze the higher level properties of the lower level system. That the mathematics governing connectionist networks can be analyzed for emergent properties seems a considerably better bet than extremely complex rule-based systems being analyzable for their emergent properties. The enterprise of analyzing the emergent properties of connectionist systems is rather closely related to traditional kinds of analysis of dynamical systems in physics; it has already shown signs that it may ultimately be as successful.
- Finally, the hard side has had priority for several decades now with disappointing results. It's time to give the soft side a few decades to produce disappointing results of its own.

The choice of adopting a fundamentally soft approach and building a hard level on top of that has serious costs—as pointed out in some detail by Kirsh in his paper in this volume. The power of symbols and symbolic computation are not given to you for free; you have to construct them out of soft stuff, and this is really very difficult. At this point, we don't know how to pull it off. As Kirsh points out, if you don't have symbols in the usual sense, it's not clear that you can cope with a number of problems. Fodor and Pylyshyn's critique is basically a statement of the same general sort: that the price one has to pay for going connectionist is the failure to account for certain regularities of the hard side, regularities that the symbolic formalism gives you essentially for free.

If the force of such critiques is taken to be that connectionism doesn't *yet* come close enough to providing the capabilities of symbolic computation to do justice to the hard side of the Paradox, then I personally think that they are quite correct. Adopting the subsymbolic stance on the Paradox amounts to taking out an enormous loan—a loan that has barely begun to be paid off.

If, on the other hand, the force of such critiques is taken to be that connectionism can *never* come close enough to providing the capabilities of symbolic computation without merely implementing the symbolic approach, then, as I will argue in the remainder of this article, I believe such critiques must be rejected.

Where are the benefits of going with the subsymbolic approach to the Paradox? Why is this large loan worth taking out? In my view, the principal justification is that if we succeed in building symbols and symbol manipulation out of "connectoplasm" then we will have an explanation of *where symbols and symbol manipulation come from*—and that is worth the risk and the effort; very much so. With any luck we'll even have an explanation how the *brain* builds symbolic computation. But even if we don't get that directly, it'll be the first theory of how to get symbols out of anything that remotely resembles the brain—and that certainly will be helpful (indeed, I would



argue, crucial) in figuring out how the brain actually does it.

Another potential payback is a way of explaining *why* those aspects of cognition that exhibit hardness should exhibit hardness: why the area of hardness falls where it does; why it is limited as it is; why the symbolic approach succeeds where it succeeds and fails where it fails.

Finally, of course, if the subsymbolic approach succeeds, we'll have a truly unified solution to the Paradox: no denial of one half of the problem, and no profoundly split brain.

We can already see contributions leading towards these ultimate results. The connectionist approach is producing new concepts and techniques for capturing the regularities in cognitive performance both at the lower level where the connectionist framework naturally applies and at the higher level where the symbolic accounts are important. (For recent surveys, see McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986; Smolensky, forthcoming). The theoretical repertoire of cognitive and computer science are being enriched by new conceptions of how computation can be done.

As far as where we actually stand on achieving the ultimate goals, in my opinion, what we have are interesting techniques and promising suggestions. Our current position in the intellectual history of connectionist computation, in my view, can be expressed by this analogy:

$$\frac{\text{current understanding of connectionist computation}}{\text{current understanding of symbolic computation}} :: \frac{\text{Aristotle}}{\text{Turing}}$$

We are somewhere approximating Aristotle's position in the intellectual development of this new computational approach. If there are any connectionist enthusiasts who think that we can really model cognition from such a position, they are, I fear, sadly mistaken. And if we can't get from Aristotle to (at least) Turing in our understanding of subsymbolic computation, we're not going to get much closer to real cognition than we are now.

One final comment before proceeding to Fodor and Pylyshyn's critique. The account given here relating the choice of a connectionist framework to the hard/soft paradox sheds some light on the question, often asked by observers of the sociology of connectionism: "Why does the connectionist fan club include such a strange assortment of people?" At least in the polite reading of this question, "strange assortment" refers to a philosophically quite heterogeneous group of cognitive scientists whose views have little more in common than a rejection of the mainstream symbolic paradigm. My answer to this question is that the priority of the hard has made a lot of people very unhappy for a long time. The failure of mainstream formal accounts of cognitive processes to do justice to the soft side of the Paradox has made people from a lot of different perspectives feel alienated from the endeavor. By assigning to the soft the position of priority, by making it the basis of the formalism, connectionism has given a lot of people who haven't had a formal leg to stand on a formal leg to stand on. And they *should* be happy about that.

At this point, "connectionism" refers more to a formalism than a theory. So it's not appropriate to paraphrase the question of the previous paragraph as "What kind of theory would have as its adherents such a disparate group of people?" It's not really a question of a *theory* at all—it's really a question of what kind of *formalism* allows people with different theories to say what they need to say.

Having made my case that understanding the choice of a connectionist formalism involves considering alternative stances towards the Paradox of Cognition, I now proceed to consider Fodor and Pylyshyn's critique in this light.

## 2. Fodor and Pylyshyn on the constituent structure of mental states

Here is a quick summary of the central argument of Fodor & Pylyshyn (1988).

- (1) Thoughts have composite structure.

By this they mean things like: the thought that *John loves the girl* is not atomic; it's a composite mental state built out of thoughts about *John*, *loves*, and *the girl*.

- (2) Mental processes are sensitive to this composite structure.

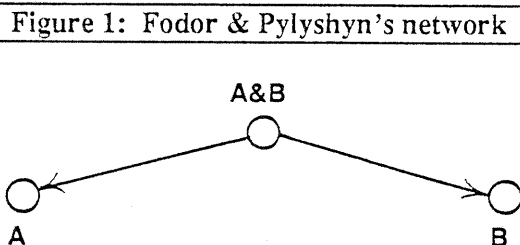
For example, from any thought of the form  $p \ \& \ q$ —regardless of what  $p$  and  $q$  are—we can deduce  $p$ .

Fodor and Pylyshyn elevate (1) and (2) to the status of defining the Classical View of Cognition, and they want to say that this is what is being challenged by the connectionists. I will later argue that they're wrong, but now we continue with their argument.

Having identified claims (1) and (2) as definitive of the Classical View, Fodor and Pylyshyn go on to argue that there are compelling arguments for these claims. [They admit up front that these arguments are a rerun updated for the 80's, a colorized version of a film that was shown in black and white some time ago—with the word "behaviorism" replaced throughout by "connectionism."] Mental states have, according to these arguments, the properties of productivity, systematicity, compositionality, and inferential coherence. Without going into all these arguments, let me simply state that for present purposes I'm willing to accept that they are convincing enough to justify the conclusion that (1) and (2) must be taken quite seriously. Whatever the inclinations of other connectionists, these and related arguments convince me that denying the hard is a mistake. They don't convince me that I should deny the soft—nor, presumably, are they intended to.

Now for Fodor and Pylyshyn's analysis of connectionism. They assert that in (standard) connectionism, *all representations are atomic*; mental states have no composite structure, violating (1). Furthermore, they assert, (standard) *connectionist processing is association* which is sensitive only to *statistics*, not to *structure*—in violation of (2). Therefore, they conclude, (standard) connectionism is maximally non-Classical: it violates both the defining principles. Therefore connectionism is defeated by the compelling arguments in favor of the Classical View.

What makes Fodor and Pylyshyn say that connectionist representations are atomic? The second figure (p. 16) of their paper says it all—it is rendered here as Figure 1. This network is supposed to illustrate the standard connectionist account of the inference from  $A \ \& \ B$  to  $A$  and to  $B$ . It is true that Ballard and Hayes wrote a paper (Ballard & Hayes, 1984) about using connectionist networks to do resolution theorem proving in which networks like this appear. However it is a serious mistake to view this as the paradigmatic connectionist account for anything like human inferences of this sort. This kind of *ultra-local* connectionist representation, in which entire propositions are represented by individual nodes, is far from typical of connectionist models, and certainly not to be taken as *definitive* of the connectionist approach.



My central counter-argument to Fodor and Pylyshyn starts with the claim that any critique of the connectionist approach must consider the consequences of using *distributed representations*, in which the representation of high level conceptual entities such as propositions are distributed over many nodes, and the same nodes simultaneously participate in the representation of many entities. Their response, in Section 2.1.3, (p. 19) is as follows. The distributed/local representation issue concerns (they assume) whether each of the nodes in Figure 1 refers to something complicated and lower level (the distributed case) or not (the local case). But, they claim, this issue is irrelevant, because it pertains to a *between level* issue, and the compositionality of mental states is a *within level* issue.

My response is that they are correct that compositionality is a within level issue, and correct that the distributed/local distinction is a between level issue. Their argument presumes that because of this difference, one issue cannot influence the other. But that is a fallacy. It assumes that the between-level relation in distributed representations can not have any consequences on the *within level* structure of the relationships between the representations of *A & B* and the representation of *A*. And that's simply false. There are implications of distributed representations for compositionality, which I'm going to bring out in the rest of this section through an extended example. In particular it will turn out that Figure 1 is no more relevant to a distributed connectionist account of inference than it is to a symbolic account. In the hyper-local case, Figure 1 is relevant and their critique stands; in the distributed case, Figure 1 is a bogus characterization of the connectionist account and their critique completely misses its target. It will further turn out that a valid analysis of the actual distributed case, based on suggestions of Pylyshyn himself, leads to quite the opposite conclusion: connectionist models using distributed representations describe mental states with a relevant kind of (within level) constituent structure.

Before developing this counter-argument, let me summarize the bottom line of the Fodor and Pylyshyn paper. Since they believe *standard* connectionism to be fatally flawed, they advocate that connectionists pursue instead a *nonstandard* connectionism. Connectionists should embrace principles (1) & (2); they should accept the classical view and should design their nets to be implementations of classical architectures. The logic implicit here is that connectionist models that respect (1) and (2) must necessarily be implementations of a classical architecture; this is their second major fallacy, which I will return to in Section 3. Fodor and Pylyshyn claim that connectionism should be used to implement classical architectures, and that having done this, connectionism will provide not a new cognitive architecture but an implementation for the old cognitive architecture—that what connectionism can provide therefore is not a new paradigm for cognitive science but rather some new information about "implementation science" or, possibly, neuroscience.

If connectionists were to follow the implementation strategy that Fodor and Pylyshyn advocate, I do believe these consequences concerning cognitive architecture *would* indeed follow. But I do not believe that it follows from accepting (1) and (2) that connectionist networks must be implementations. In Section 3, I argue that connectionists can consistently accept (1) and (2) while rejecting the implementationalist approach Fodor and Pylyshyn advocate.

For now, the goal is to show that connectionist models using *distributed* representations ascribe to mental states the kind of compositional structure demanded by (1), contrary to Fodor and Pylyshyn's conclusion based on the network of Figure 1 embodying a hyper-local representation.

My argument consists primarily in carrying out an analysis that was suggested by Zenon Pylyshyn himself at the 1984 Cognitive Science Meeting in Boulder. A sort of debate about connectionism was held between Geoffrey Hinton and David Rumelhart on the one hand, and Zenon Pylyshyn and Kurt VanLehn on the other. While pursuing the nature of connectionist representations, Pylyshyn asked Rumelhart: "Look, can you guys represent a cup of coffee in these networks?" Rumelhart's reply was "Sure" so Pylyshyn continued: "And can you represent a cup without coffee in it?" Waiting for the trap to close, Rumelhart said "Yes" at which point Pylyshyn pounced: "Ah-hah, well, the difference between the two is just the representation of *coffee* and you've just built a representation of *cup with coffee* by combining a representation of *cup* with a representation of *coffee*."

So let's carry out exactly the construction suggested by Pylyshyn, and see what conclusion it leads us to. We'll take a *distributed* representation of *cup with coffee* and subtract from it a distributed representation of *cup without coffee* and we'll call what's left "the connectionist representation of *coffee*."

To generate these distributed representations I will use a set of "microfeatures" (Hinton, McClelland, & Rumelhart, 1986) that are not very micro—but that's always what happens when you try to create examples that can be intuitively understood in a nontechnical exposition. These microfeatures are shown in Figure 2.

Figure 2: Representation of *cup with coffee*

Units	Microfeatures
●	upright container
●	hot liquid
○	glass contacting wood
●	porcelain curved surface
●	burnt odor
●	brown liquid contacting porcelain
●	porcelain curved surface
○	oblong silver object
●	finger-sized handle
●	brown liquid with curved sides and bottom

Figure 2 shows a distributed representation of *cup with coffee*: a pattern of activity in which those units that are active (black) are those that correspond to microfeatures present in the description of a cup containing coffee. Obviously, this is a crude, nearly sensory-level representation, but again that helps make the example more intuitive—it's not essential.

Given the representation of *cup with coffee* displayed in Figure 2, Pylyshyn suggests we subtract the representation of *cup without coffee*. The representation of *cup without coffee* is shown in Figure 3, and Figure 4 shows the result of subtracting it from the representation of *cup with coffee*.

Figure 3: Representation of *cup without coffee*

Units	Microfeatures
●	upright container
○	hot liquid
○	glass contacting wood
●	porcelain curved surface
○	burnt odor
○	brown liquid contacting porcelain
●	porcelain curved surface
○	oblong silver object
●	finger-sized handle
○	brown liquid with curved sides and bottom

Figure 4: "Representation of *coffee*"

Units	Microfeatures
○	upright container
●	hot liquid
○	glass contacting wood
○	porcelain curved surface
●	burnt odor
●	brown liquid contacting porcelain
○	porcelain curved surface
○	oblong silver object
○	finger-sized handle
●	brown liquid with curved sides and bottom

So what does this procedure produce as "the connectionist representation of *coffee*"? Reading off from Figure 4, we have a burnt odor and hot brown liquid with curved sides and bottom surfaces contacting porcelain. This is indeed a representation of *coffee*, but in a very particular context: the context provided by *cup*.

What does this mean for Pylyshyn's conclusion that "the connectionist representation of *cup with coffee* is just the representation of *cup without coffee* combined with the representation of *coffee*"? What is involved in combining the representations of Figures 3 and 4 back together to form that of Figure 2? We assemble the representation of *cup with coffee* from a representation of a *cup*, and a representation of *coffee*, but it's a rather strange combination. It's also got representation of the *interaction* of the cup with coffee—like *brown liquid contacting porcelain*. Thus the composite representation is built from *coffee extracted* from the situation *cup with coffee*, together with *cup extracted* from the situation *cup with coffee*, together with their interaction.

So the compositional structure is there, but it's there in an *approximate* sense. It's *not* equivalent to taking a context-independent representation of *coffee* and a context-independent representation of *cup*—and certainly not equivalent to taking a context-independent representation of the relationship *in* or *with*—and sticking them all together in a symbolic structure, concatenating them together to form the kinds of syntactic compositional structures that Fodor and Pylyshyn think connectionist nets should implement.

To draw this point out further, let's reconsider the representation of *coffee* once the cup has been subtracted off. This, suggests Pylyshyn, is the connectionist representation of *coffee*. But as we have already observed, this is really a representation of *coffee* in the particular context of being inside a cup. According to Pylyshyn's formula, to get the connectionist representation of *coffee* it should have been in principle possible to take the connectionist representation of *can with coffee* and subtract from it the connectionist representation of *can without coffee*. What would happen if we actually did this? We would get a representation of ground brown burnt smelling granules stacked in a cylindrical shape, together with granules contacting tin. This is the connectionist representation of *coffee* we get by starting with *can with coffee* instead of *cup with coffee*. Or we could start with the representation of *tree with coffee* and subtract off *tree without coffee*. We would get a connectionist representation for *coffee* which would be a representation of brown beans in a funny shape hanging suspended in mid air. Or again we could start with *man with coffee* and get still another connectionist representation of *coffee*: one quite similar to the entire representation of *cup with coffee* from which we extracted our first representation of *coffee*.

The point is that the representation of *coffee* that we get out of the construction starting with *cup with coffee* leads to a different representation of *coffee* than we get out of other constructions that have equivalent status a priori. That means if you want to talk about the connectionist representation of *coffee* in this distributed scheme, you have to talk about a *family of distributed activity patterns*. What knits together all these particular representations of *coffee* is nothing other than a *family resemblance*.

The first moral I want to draw out of this *coffee* story is this: unlike the hyper-local case of Figure 1, with distributed representations, complex representations *are* composed of representations of constituents. The constituency relation here is a *within level* relation, as Fodor and Pylyshyn require: the pattern or *vector* representing *cup with coffee* is composed of a *vector* that can be identified as a distributed representation of *cup without coffee* together with a *vector* that can be identified as a particular distributed representation of *coffee*. In characterizing the constituent vectors of the vector representing the composite, we are *not* concerned with the fact that the vector representing *cup with coffee* is a vector comprised of the activity of individual microfeature units. The *between level* relation between the vector and its individual numerical elements is *not* the constituency relation, and so section 2.1.4 (p. 19–28) of Fodor & Pylyshyn (1988) is irrelevant—there they address a mistake that is not being made.

The second moral is that the constituency relation among distributed representations is one that is important for the analysis connectionist models, and for explaining their behavior, but it is *not* a part of the causal mechanism within the model. In order to process the vector representing *cup with coffee*, the network does not have to decompose it into constituents. For processing, it is the *between level* relation, not the within level relation, that matters. The processing of the vector representing *cup with coffee* is determined by the individual numerical activities that make up the vector: it is over these lower-level activities that the processes are defined. Thus the fact that there is considerable arbitrariness in the way the constituents of *cup with coffee* are defined introduces no ambiguities in the way the network processes that representation—the ambiguities exist only for us who analyze the model and try to explain its behavior. Any particular definition of constituency that gives us explanatory leverage is a valid definition of constituency; lack of uniqueness is not a problem.

This leads directly to the third moral, that the decomposition of composite states into their constituents is not precise and uniquely defined. The notion of constituency is important but attempts to formalize it are likely to crucially involve *approximation*. As discussed at some length in Smolensky (1988a), this is the typical case: notions from symbolic computation provide important tools for constructing higher-level accounts of the behavior of connectionist models using distributed representation—but these notions provide approximate, not precise, accounts.

Which leads to the fourth moral, that while connectionist networks using distributed representations *do* describe mental states with the type of constituency required by (1), they do *not* provide a literal implementation of a syntactic language of thought. The context dependency of the constituents, the interactions that must be accommodated when they are combined, the inability to uniquely, precisely identify constituents, the need to take seriously the notion that the representation of *coffee* is a collection of vectors knit together by family resemblance—all these entail that the relation between connectionist constituency and syntactic constituency is *not* one of literal implementation. In particular, it would be absurd to claim that even if the connectionist story is correct then that would have no implications for the cognitive architecture, that it would merely fill in lower level details without important implications for the higher level account.

These conclusions all address (1) without explicitly addressing (2). Addressing (2) properly is far beyond the scope of this paper. To a considerable extent, it is beyond the scope of current connectionism. Let me simply point out that the Structure/Statistics Dilemma has an attractive possible solution that the connectionist approach is perfectly situated to pursue: *the mind is an statistics-sensitive engine operating on structure-sensitive numerical representations*. The previous arguments have shown that distributed representations do possess constituency relations, and that, properly analyzed, these representations can be seen to encode structure. Extending this to grapple with the full complexity of the kinds of rich structures implicated in complex cognitive processes is a research problem that has been attacked with some success but which remains to be definitively concluded (see Smolensky, 1987, and Section 3 below). Once we have complex structured information represented in distributed

numerical patterns, statistics sensitive processes can proceed to analyze the statistical regularities in a fully structure-sensitive way. Whether such processes can cope with the full force of the Structure/Statistics Dilemma is apt to remain an open question for some time yet.

The conclusion, then, is that distributed models *can* satisfy both (1) and (2). Whether (1) and (2) can be satisfied to the point of providing an account adequate to cover the *full demands of cognitive modeling* is of course an open empirical question—just as it is for the symbolic approach to satisfying (1) and (2). Just the same, distributed connectionist models do *not* amount to an implementation of the symbolic instantiations of (1) and (2) that Fodor and Pylyshyn are committed to.

Before summing up, I'd like to return to Figure 1. In what sense can Figure 1 be said to describe the relation between the distributed representation of A&B and the distributed representations of A and B? It was the intent of the *coffee* example to show that the distributed representations of the constituents are, in an approximate but explanation-relevant sense, part of the representation of the composite. Thus, in the distributed case, the relation between the node of Figure 1 labelled A&B and the others is a sort of whole/part relation. An inference mechanism that takes as input the vector representing A&B and produces as output the vector representing A is a mechanism that extracts a part from a whole. And in this sense it is no different from a symbolic inference mechanism that takes the syntactic structure **A & B** and extracts from it the syntactic constituent **A**. The connectionist mechanisms for doing this are of course quite different than the symbolic mechanisms, and the approximate nature of the whole/part relation gives the connectionist computation different overall characteristics: we don't have simply a new implementation of the old computation.

It is clear that, just as Figure 1 offers a crude summary of the symbolic process of passing from **A & B** to **A**, a summary that uses the labels to encode hidden internal structures within the nodes, *exactly the same is true of the distributed connectionist case*. In the distributed case, just as in the symbolic case, the links in Figure 1 are crude summaries of complex processes and not simple-minded causal channels that pass activity from the top node to the lower nodes. Such a causal story applies only to the hyper-local connectionist case, which here serves as the proverbial straw man.

Let me be clear: there is no distributed connectionist model, as far as I know, of the kind of formal inference Fodor and Pylyshyn have in mind here. Such formal inference is located at the far extreme of the hard side of the Paradox, and is not at this point a cognitive process (or abstraction thereof) that the connectionist formalism can be said to have built upon its soft substrate. But at root the Fodor and Pylyshyn critique revolves around the constituent structure of mental states—formal inference is just one setting in which to see the importance of that constituent structure. So the preceding discussion of the constituent structure of distributed representations does address the heart of their critique, even if a well-developed connectionist account of formal inference remains unavailable.

So let's summarize the overall picture at this point. We've got principles (1) and (2), and we've got a symbolic instantiation of these in a language of thought using syntactic constituency. According to Fodor and Pylyshyn, what connectionists should do is take that symbolic language of thought as a higher level description and then produce a connectionist implementation in a literal sense. The syntactic operations of the symbolic language of thought then provide an exact formal higher level account.

By contrast, I argue that the distributed view of connectionist compositionality allows us to instantiate the same basic principles of (1) and (2) *without* going through a symbolic language of thought. By going straight to distributed connectionist models we get *new instantiations of compositionality principles*.

I happen to believe that the symbolic descriptions *do* provide useful approximate higher level accounts of how these connectionist models compute—but in no sense do these distributed connectionist models provide a literal implementation of a symbolic language of thought. The approximations require a willingness to accept context sensitive symbols and interactional components present in compositional structures, and the other funny business that came out in the *coffee* example. If you're willing to live with all those degrees of approximation then you can usefully view these symbolic level descriptions as approximate higher level accounts of the processing in a connectionist network.

The overall conclusion, then, is that *the classical and connectionist approaches differ not in whether they accept principles (1) and (2), but in how they formally instantiate them.* To confront the real classical/connectionist dispute, one has to be willing to descend to the level of the particular formal instantiations they give to these nonformal principles. To fail to descend to this level of detail is to miss the issue. In the classical approach, principles (1) and (2) are formalized using syntactic structures for thoughts and symbol manipulation for mental processes. In the connectionist view (1) or (2) are formalized using distributed vectorial representations for mental states, and the corresponding notion of compositionality, together with association-based mental processes that derive their structure sensitivity from the structure sensitivity of the vectorial representations engaging in those processes.

In terms of research methodology, this means that the agenda for connectionism should be not be to develop a connectionist implementation of the symbolic language of thought but rather to develop formal analysis of vectorial representations of complex structures and operations on those structures that are sufficiently structure-sensitive to do the required work.

In summary: distributed representations provide a description of mental states with semantically interpretable constituents, but there is no precise formal account of the construction of composites from context-independent semantically interpretable constituents. On this account, there *is* a language of thought—but only approximately; the language of thought does not provide a basis for an exact formal account of mental structure or processes—it cannot provide a precise formal account of the cognitive architecture.<sup>3</sup>

### 3. Connectionism and implementation

In Section 2 I argued that connectionist research should be directed toward structure-sensitive representations and processes but not toward the implementation of a symbolic language of thought. In this section I want to consider this middle ground between implementing symbolic computation and ignoring structure. Many critics of connectionism do not seem to understand that this middle ground exists. (For further discussion of this point, and a map that explicitly locates this middle ground, see Smolensky, 1988b.)

A rather specific conclusion of Section 2 was that connectionists need to develop the analysis of distributed (vectorial) representations of composite structures and the kinds of processes that operate on them with the necessary structure sensitivity. More generally, my characterization of the goal of connectionist modeling is to develop formal models of cognitive processes that are based on the mathematics of dynamical systems continuously evolving in time: complex systems of numerical variables governed by differential equations. These formal accounts live in the category of continuous mathematics rather than relying on the discrete mathematics that underlies the traditional symbolic formalism. This characterization of the goal of connectionism is far from universal: it's quite inconsistent with the definitive characterization of Feldman & Ballard (1982), for example. In Smolensky (1988a) I argue at some length that my characterization, called *PTC*, constitutes a Proper Treatment of Connectionism.

A central component of *PTC* is the relation hypothesized between connectionist models based on continuous mathematics and classical models based on discrete, symbolic computation. That relationship, which entered briefly in the Fodor and Pylyshyn argument of Section 2, might be called the *cognitive correspondence principle*: When connectionist computational systems are analyzed at higher levels, elements of symbolic computation appear as emergent properties.

Figure 5 illustrates the cognitive correspondence principle. At the top we have nonformal notions: the central hypotheses that the principles of cognition consist in principles of memory, of inference, of compositionality and constituent structure, etc. In the Fodor and Pylyshyn argument, the relevant nonformal principles were their compositionality principles (1) and (2).





without being prejudiced by archaic prescientific notions such as those at the top of Figure 5.

In rejecting both the implementationalist and eliminativist positions, PTC views connectionist accounts as reducing and explaining symbolic accounts. Connectionist accounts serve to refine symbolic accounts, to reduce the degree of approximation required, to enrich the computational notions from the symbolic and discrete world, to fill them out with notions of continuous computation. Primarily that's done by descending to a lower level of analysis, by focussing on the microstructure implicit in these kinds of symbolic operations.

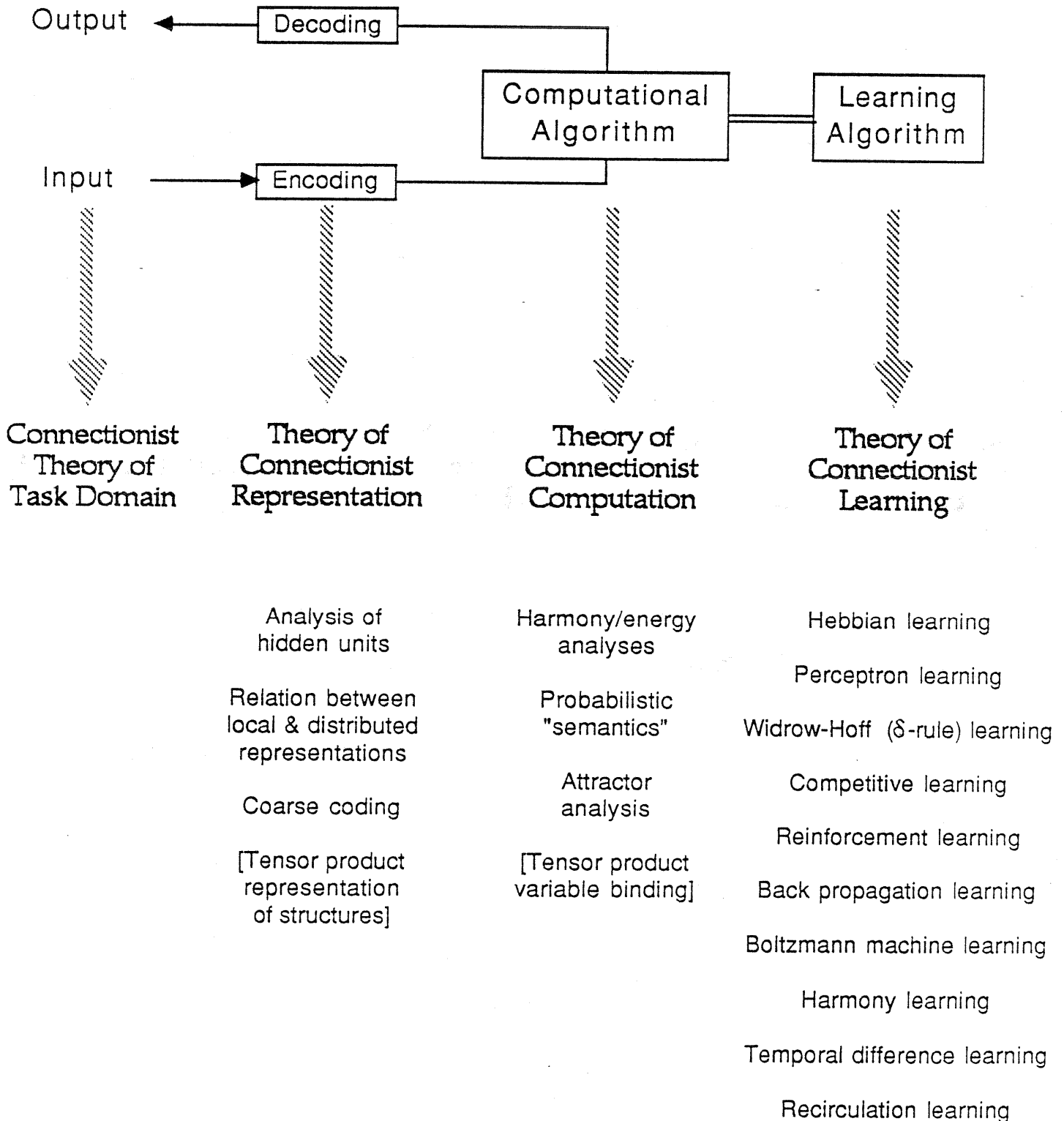
I call this the cognitive correspondence principle because I believe it has a role to play in the developing microtheory of cognition that's analogous to the role that the quantum correspondence principle played in the development of microtheory in physics. The case from physics embodies the structure of Figure 5 quite directly. There are certain physical principles that arch over both the classical and quantum formalisms: the notions of space and time and associated invariance principles, the principles of energy and momentum conservation, force laws, and so on. These principles at the top of Figure 5 are instantiated in particular ways in the classical formalism, corresponding to the point one level down on the right branch. To go to a lower level of physical analysis requires the development of a new formalism. In this quantum formalism, the fundamental principles are reinstated: they occupy the bottom of the left branch. The classical formalism can be looked at as a higher level description of the same principles operating at the lower quantum level: the dotted line of Figure 5. Of course quantum mechanics does not *implement* classical mechanics: the accounts are intimately related, but classical mechanics provides an approximate, not an exact, higher-level account.<sup>4</sup> In a deep sense, the quantum and classical theories are quite incompatible: according to the ontology of quantum mechanics, the ontology of classical mechanics is quite impossible to realize in this world. But there is no denying that the classical ontology and the accompanying principles are theoretically essential, for at least two reasons: (a) to provide explanations (in a literal sense, approximate ones) of an enormous range of classical phenomena for which direct explanation from quantum principles is hopelessly infeasible, and (b), historically, to provide the guidance necessary to discover the quantum principles in the first place. To try to develop lower level principles without looking at the higher level principles for guidance, given the insights we have gained from those principles, would seem, to put it mildly, inadvisable. It is basically this pragmatic consideration that motivates the cognitive correspondence principle and the PTC position it leads to.

In the PTC methodology, it is essential be able to analyze the higher level properties of connectionist computation in order to relate them to properties of symbolic computation, e.g., to see whether they have the necessary computational power. I now want to summarize what I take to be the state of the art in the mathematical analysis of computation in connectionist systems, and how it relates to Fodor and Pylyshyn's critique. This summary is presented in Figure 6.

Figure 6 shows the pieces of a connectionist model and elements of their analysis. The connectionist model basically has four parts. There's the task that the model is supposed to perform—for example, to take some set of inputs into a set of outputs described in the terms characteristic of the problem-domain. Then there is an actual connectionist network which will perform that mapping from input to output; but between the original task and the model we need methods for encoding and decoding. The encoding must take the problem-domain characterization of an input and code it into a form that the network can process, namely, activities of certain input processors. Similarly, the activity of the output processors has to be decoded into some problem domain statement which can be construed as the output of the network. The input-to-output mapping inside the network is the computational algorithm embodied in the network and, more often than not, in addition, there's a learning algorithm which modifies the parameters in the computational algorithm in order to get it to converge on the correct input/output behavior of the correct computation.

In terms of analyzing these four elements of connectionist modeling, things get progressively worse as we move from right to left. In the area of connectionist learning, there are lots of analyses: algorithms for tweaking lower-level connection strengths which will produce reasonable higher level convergence towards the correct input/output mapping. The figure shows as many as would fit conveniently and there are many more.<sup>5</sup>

**Figure 6: Theory of Connectionist Models**





So, if you think that the problem with connectionism is that a particular learning algorithm has some characteristic you don't like, then chances are there is another learning algorithm that will make you happy. Relative to the rest, the learning theory is in good shape, even though when it comes to theorems about what functions can be learned by a given algorithm, there's very little.

With respect to analyzing the higher-level properties of the algorithms for computing outputs from inputs, there's considerably less theory. The technique of analyzing convergence using a function that measures the "energy" or "harmony" of network states (Ackley, Hinton, & Sejnowski, 1985; Cohen & Grossberg, 1983; Geman & Geman, 1984; Hinton & Sejnowski, 1983; Hopfield, 1982; Smolensky, 1983, 1986a) get us somewhere, as do a few other techniques<sup>6</sup> but it seems rather clear that the state of analysis of connectionist computation is considerably less developed than that of connectionist learning.

After this things get *very* thin. What about the theory behind encoding and decoding, the theory of how to take the kinds of inputs and outputs that have to be represented for cognitive processes and turn them into actual patterns of activity? By and large, it's a black art: there is not much in the way of analysis. People have been getting their hands dirty exploring the representations in hidden units (e.g. Hinton, 1986; Rosenberg, 1987), but so far I see little reason to believe our understanding of these representations will go further than understanding an occasional node or a few statistical properties. There are a few other simple analyses<sup>7</sup> but they don't take us very far.

At the far left of Figure 6 is the theory of the task environment that comes out of a connectionist perspective. This is essentially nonexistent. To many, I believe, that's really the ultimate goal: the theory of the domain in connectionist terms.

As Figure 6 makes clear, there is a very important weak leg here: the connectionist theory of representation. In particular, until recently we haven't had any systematic ideas about how to represent complex structures. In fact, it was Fodor and Pylyshyn who really got me thinking about this, and ultimately convinced me. The result was the tensor product technique for generating fully distributed representations of complex structures (Smolensky, 1987). For this reason the tensor product representation is dedicated to Fodor and Pylyshyn. This representational scheme is a formalization and generalization of representational techniques that have been used piecemeal in connectionist models. As others have discussed in this volume, the tensor product technique provides a systematic and disciplined procedure for representing complex, structured objects. One can prove that the tensor product representation has a number of nice computational properties from the standpoint of connectionist processing. In this sense, it is appropriate to view the tensor product representation as occupying the lower level corner of Figure 5: it provides a formalization that is natural for connectionist computation of the nonformal notion of constituent structure, and is a likely candidate to play a role in connectionist cognitive science analogous to that played by constituent structure trees in symbolic cognitive science.

The tensor product representation rests on the use of the tensor product operation to perform in the vectorial world the analog of binding together a variable and its value. Figure 6 shows where tensor product variable binding and tensor product representations of structures fit into the overall problem of analyzing connectionist cognitive models.

I hope this last section has made more plausible my working hypothesis that between the connectionist view that Fodor and Pylyshyn attack—denying the importance of structured representations and structure-sensitive processes—and the connectionist methodology they advocate—implementation of the classical symbolic cognitive architecture—there is a promising middle ground on which productive and exciting research can be pursued.

#### Acknowledgements

This work has been supported by NSF grants IRI-8609599 and ECE-8617947 to the author, and by a grant to the author from the Sloan Foundation's computational neuroscience program.

## Footnotes

1. For related discussions, see, e.g., Gerken & Bever, 1986; Greeno, 1987.
2. For some somewhat spooky empirical results directly bearing on this issue, see Bever, Carrithers, & Townsend, 1987.
3. An important open question is whether the kind of story I have given on *cup of coffee* using these hokey microfeatures will carry over to the kind of distributed representations that real connectionist networks create for themselves in their hidden units—if you make the analysis appropriately sophisticated. The resolution of this issue depends on the (as yet inscrutable) nature of these representations for realistic problems. The nature of the problem is important, for it is perfectly likely that connectionist networks will develop compositional representations in their hidden units only when this is advantageous for the problem they are trying to solve. As Fodor and Pylyshyn, and the entire Classical paradigm, argue, such compositional representations are in fact immensely useful for a broad spectrum of cognitive problems. But until such problems—which tend to be considerably more sophisticated than those usually given to connectionist networks—have been explored in some detail with connectionist models, we won't really know if hidden units will develop compositional representations (in the approximate sense discussed in this paper) when they "should."
4. Many cases analogous to "implementation" are found in physics: Newton's laws provide an "implementation" of Kepler's laws; Maxwell's theory "implements" Coulomb's law; the quantum principles of the hydrogen atom "implement" Balmer's formula.
5. Here are a smattering of references to these learning rules; rather than giving historically primary references I have cited recent easily accessible expositions that include the original citations. (In fact I have chosen papers in Rumelhart, McClelland, and the PDP Group, 1986, when possible.) For an exposition of Hebbian, perceptron, and Widrow-Hoff or delta-rule learning, see Rumelhart, Hinton, & McClelland, 1986 and Stone, 1986. For competitive learning see Grossberg, 1987, and Rumelhart & Zipser, 1986. For reinforcement learning, see Barto, Sutton, & Anderson, 1983, and Sutton, 1987. For back propagation learning see Rumelhart, Hinton, & Williams, 1986. For Boltzmann machine learning, see Hinton & Sejnowski, 1986. For harmony learning, see Smolensky, 1986a. Temporal difference learning is reported in Sutton, 1987. A simple recirculation learning algorithm is discussed in Smolensky, 1987; the idea has been under exploration by Hinton & McClelland for several years, and their first paper should appear in 1988.
6. On giving the computation in connectionist networks semantics based on statistical inference, see Shastri & Feldman, 1985; Smolensky, 1986a; Golden, 1988.
7. For some simple explorations of the relation between local and distributed representations, see Smolensky, 1986b. For some observations about the power of the distributed representational technique called "coarse coding," see Hinton, McClelland, & Rumelhart, 1986.

## References

- Ballard, D. & Hayes, P.J. (1984). Parallel logical inference. *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*. Rochester, N.Y. June.
- Barto, A. G., Sutton, R. S., Anderson, C. W. (1983). Neuronlike elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13, 834-846.
- Bever, T. G., Carrithers, C., & Townsend, D. J. (1987). A tale of two brains: The sinistral quasimodularity of language. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, 764-773. Seattle, WA. July.
- Ackley, D.H., Hinton, G.E., & Sejnowski, T.J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147-169.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cohen, M. A. & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, & Cybernetics* SMC-13, 815-826.
- Dreyfus, S.E., & Dreyfus, H.L. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.
- Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Feldman, J.A. & Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science* 6, 205-254.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Gerken, L. & Bever, T. G. (1986). Linguistic intuitions are the result of interactions between perceptual processes and linguistic universals. *Cognitive Science* 10, 457-476.
- Golden, R. (1988). A unified framework for connectionist systems. *Biological Cybernetics*. (in press).
- Greeno, J. G. (1987). The cognition connection. *The New York Times*, Jan. 4, p. 28.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science* 11, 23-63.
- Gupta, M., Ragade, R., & Yager R. (Eds.). (1979). *Advances in fuzzy set theory and applications*. Amsterdam: North-Holland.
- Hinton, G. E. (1987). Learning distributed representations of concepts. *Proceedings of the Eighth Annual Meeting of the Cognitive Science Society*, 1-12.
- Hinton, G.E., McClelland, J.L., & Rumelhart, D.E. (1986). Distributed representations. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure*

- of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press/Bradford Books.
- Hinton, G.E. & Sejnowski, T.J. (1983a). Analyzing cooperative computation. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society.* Rochester, NY.
- Hinton, G.E., & Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA 79*, 2554–2558.
- McClelland, J. L., D. E. Rumelhart, & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, MA: MIT Press/Bradford Books.
- Rosenberg, C. R. (1987). Revealing the structure of NETtalk's internal representations. *Proceedings of the Ninth Annual Meeting of the Cognitive Science Society*, 537–554. Seattle, WA. July.
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Rumelhart, D.E., Hinton, G.E., & McClelland, J.L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Rumelhart, D. E., J. L. McClelland, & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Shastri, L. & Feldman, J.A. (1985). Evidential reasoning in semantic networks: A formal theory. *Proceedings of the International Joint Conference on Artificial Intelligence.* Los Angeles, CA.
- Shortliffe, E. H. (1976). *Computer-based medical consultations: MYCIN.* New York: American Elsevier.
- Smolensky, P. (1983). Schema selection and stochastic inference in modular environments. *Proceedings of the National Conference on Artificial Intelligence.* Washington, DC.
- Smolensky, P. (1986a). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.
- Smolensky, P. (1986b). Neural and conceptual interpretations of parallel distributed processing models. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, *Parallel Distributed Processing: Explorations*



- in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, MA: MIT Press/Bradford Books.
- Smolensky, P. (1987). On variable binding and the representation of symbolic structures in connectionist systems. Technical Report CU-CS-355-87, Department of Computer Science, University of Colorado at Boulder, February. (Revised version to appear in *Artificial Intelligence*).
- Smolensky, P. (1988a). On the proper treatment of connectionism. *The Behavioral and Brain Sciences*. 11 (1). In press.
- Smolensky, P. (1988b). Putting together connectionism—again. *The Behavioral and Brain Sciences*. 11 (1). In press.
- Smolensky, P. (forthcoming). *Lectures on connectionist cognitive modeling*. Erlbaum.
- Stone, G.O. (1986). An analysis of the delta rule and learning statistical associations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books.
- Sutton, R. S. (1987). Learning to predict by the methods of temporal differences. Technical Report 87-509.1, GTE Laboratories, Waltham, MA.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8, 338-353.
- Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese* 30, 407-428.
- Zadeh, L. A. (1983). Role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems* 11, 199-227.