Cognition, 35 (1990) 183–204

Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work

JERRY FODOR*

Rutgers University and CUNY Graduate Center

BRIAN P. McLAUGHLIN

Rutgers University

Received May 23, 1989, final revision accepted August 18, 1989

Abstract

Fodor, J., and McLaughlin, B.P., 1990. Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. Cognition, 35. 183–204.

In two recent papers, Paul Smolensky responds to a challenge Jerry Fodor and Zenon Pylyshyn posed for connectionist theories of cognition: to explain the existence of systematic relations among cognitive capacities without assuming that mental processes are causally sensitive to the constituent structure of mental representations. Smolensky thinks connectionists can explain systematicity if they avail themselves of "distributed" mental representations. In fact, Smolensky offers two accounts of distributed mental representation, corresponding to his notions of "weak" and "strong" compositional structure. We argue that weak compositional structure is irrelevant to the systematicity problem and of dubious internal coherence. We then argue that strong compositional (tensor product) representations fail to explain systematicity because they fail to exhibit the sort of constituents that can provide domains for structure sensitive mental processes.

Introduction

In two recent papers, Paul Smolensky (1987, 1988b) responds to a challenge Jerry Fodor and Zenon Pylyshyn (Fodor & Pylyshyn, 1988) have posed for connectionist theories of cognition: to explain the existence of systematic relations among cognitive capacities without assuming that cognitive proces-

*Authors are listed alphabetically. Requests for reprints should be addressed to Jerry Fodor, Department of Philosophy, The Graduate Center, CUNY, 33 West 42nd Street, New York, NY 10036, U.S.A.

ses are causally sensitive to the constituent structure of mental representations. This challenge implies a dilemma: if connectionism can't account for systematicity, it thereby fails to provide an adequate basis for a theory of cognition; but if its account of systematicity requires mental processes that are sensitive to the constituent structure of mental representations, then the theory of cognition it offers will be, at best, an implementation architecture for a "classical" (language of thought) model. Smolensky thinks connectionists can steer between the horns of this dilemma if they avail themselves of certain kinds of distributed mental representation. In what follows, we will examine this proposal.

Our discussion has three parts. In section I, we briefly outline the phenomenon of systematicity and its Classical explanation. As we will see, Smolensky actually offers two alternatives to this Classical treatment, corresponding to two ways in which complex mental representations can be distributed; the first kind of distribution yields complex mental representations with "weak compositional structure", the second yields mental representations with "strong compositional structure". We will consider these two notions of distribution in turn: in section II, we argue that Smolensky's proposal that complex mental representations have weak compositional structure should be rejected both as inadequate to explain systematicity and on internal grounds; in section III, we argue that postulating mental representations with strong compositional structure also fails to provide for an explanation of systematicity. The upshot will be that Smolensky avoids only one horn of the dilemma that Fodor and Pylyshyn proposed. We shall see that his architecture is genuinely non-Classical since the representations he postulates are not "distributed over" constituents in the sense that Classical representations are; and we shall see that for that very reason Smolensky's architecture leaves systematicity unexplained.

I. The systematicity problem and its Classical solution

The systematicity problem is that cognitive capacities come in clumps. For example, it appears that there are families of semantically related mental states such that, as a matter of psychological law, an organism is able to be in one of the states belonging to the family only if it is able to be in many of the others. Thus, you don't find organisms that can learn to prefer the green triangle to the red square but can't learn to prefer the red triangle to the green square. You don't find organisms that can think the thought that the girl loves John but can't think the thought that John loves the girl. You don't find organisms that can infer P from P&Q&R but can't infer P from P&Q.

And so on over a very wide range of cases. For the purposes of this paper, we assume without argument:

(i) that cognitive capacities are generally systematic in this sense, both in humans and in many infrahuman organisms;

(ii) that it is nomologically necessary (hence counterfactual supporting) that this is so;

(iii) that there must therefore be some psychological mechanism in virtue of the functioning of which cognitive capacities are systematic;

(iv) and that an adequate theory of cognitive architecture should exhibit this mechanism.

Any of i–iv may be viewed as tendentious; but, so far as we can tell, all four are accepted by Smolensky. So we will take them to be common ground in what follows.[1]

The Classical account of the mechanism of systematicity depends crucially on the idea that mental representation is language-like. In particular, mental representations have a combinatorial syntax and semantics. We turn to a brief discussion of the Classical picture of the syntax and semantics of mental representations; this provides the basis for understanding the Classical treatment of systematicity.

---

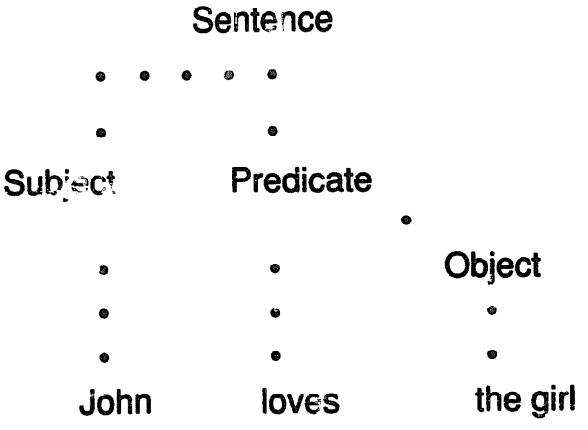
[1]Since the two are often confused, we wish to emphasize that taking *systematicity* for granted leaves the question of *compositionality* wide open. The systematicity of cognition consists of, for example, the fact that organisms that can think aRb can think bRa and vice versa. *Compositionality* proposes a certain explanation of systematicity: viz., that the content of thoughts is determined, in a uniform way, by the content of the context-independent concepts that are their constituents; and that the thought that bRa is constituted of the same concepts as the thought that aRb. So the polemical situation is as follows. If you are a Connectionist who accepts systematicity, then you must argue either that systematicity can be explained without compositionality, or that connectionist architecture accommodates compositional representation. So far as we can tell, Smolensky vacillates between these options; what he calls 'weak compositionality" favors the former and what he calls "strong compositionality" favors the latter.

We emphasize this distinction between systematicity and compositionality in light of some remarks by an anonymous *Cognition* reviewer: "By berating the [connectionist] modelers for their inability to represent the common-sense [uncontextualized] notion of 'coffee' ... Fodor and McLaughlin are missing a key point – the models are not supposed to do so. If you buy the ... massive context-sensitivity ... that connectionists believe in." Our strategy is *not*, however, to argue that there is something wrong with connectionism because it fails to offer an uncontextualized notion of mental (or, mutatis mutandis, linguistic) representation. Our argument is that if connectionists assume that mental representations are context sensitive, they will need to offer some explanation of systematicity that does not entail compositionality *and they do not have one*.

We do not, therefore, offer direct arguments for context-insensitive concepts in what follows; we are quite prepared that "coffee" should have a meaning only in context. Only, we argue, *if* it does, then some non-compositional account of the systematicity of coffee-throughts will have to be provided.

Classical syntax and Classical constituents

The Classical view holds that the syntax of mental representations is like the syntax of natural language sentences in the following respect: both include complex symbols (bracketing trees) which are constructed out of what we will call *Classical constituents*. Thus, for example, the English sentence "John loves the girl" is a complex symbol whose decomposition into Classical constituents is exhibited by some such bracketing tree as:



Correspondingly, it is assumed that the mental representation that is entertained when one thinks the thought that John loves the girl is a complex symbol of which the Classical constituents include representations of John, the girl, and loving.

It will become clear in section III that it is a major issue whether the sort of complex mental representations that are postulated in Smolensky's theory have constituent structure. We do not wish to see this issue degenerate into a terminological wrangle. We therefore stipulate that, for a pair of expression types E1, E2, the first is a *Classical* constituent of the second *only if* the first is tokened whenever the second is tokened. For example, the English word "John" is a Classical constituent of the English sentence "John loves the girl" and every tokening of the latter implies a tokening of the former (specifically, every token of the latter *contains* a token of the former; you can't say "John loves the girl" without saying "John").[2] Likewise, it is assumed that a men-

[2]Though we shall generally consider examples where complex symbols literally *contain* their Classical constituents, the present condition means to leave it open that symbols may have Classical constituents that are not among their (spatio-temporal) parts. (For example, so far as this condition is concerned, it might be that the Classical constituents of a symbol include the values of a "fetch" operation that takes the symbol as an argument.)

talese symbol which names John is a Classical constituent of the mentalese symbol that means that John loves the girl. So again tokenings of the one symbol require tokenings of the other.

It is precisely because Classical constituents have this property that they are always accessible to operations that are defined over the complex symbols that contain them; in particular, it is precisely because Classical mental representations have Classical constituents that they provide domains for structure-sensitive mental processes. We shall see presently that what Smolensky offers as the "constituents" of connectionist mental representations are non-Classical in this respect, and that that is why his theory provides no account of systematicity.

Classical semantics

It is part of the Classical picture, both for mental representation and for representation in natural languages, that generally when a complex formula (e.g., a sentence) S expresses the proposition P, S's constituents express (or refer to) the elements of P.[3] For example, the proposition that John loves the girl contains as its elements the individuals John and the girl, and the two-place relation "loving". Correspondingly, the formula "John loves the girl", which English uses to express this proposition, contains as constituents the expressions "John", "loves" and "the girl". The sentence "John left and the girl wept", whose constituents include the formulas "John left" and "the girl wept", expresses the proposition that John left and the girl wept, whose elements include the proposition that John left and the proposition that the girl wept. And so on.

These assumptions about the syntax and semantics of mental representations are summarized by condition C:

C: If a proposition P can be expressed in a system of mental representation M, then M contains some complex mental representation (a "mental sentence") S, such that S expresses P and the (Classical) constituents of S express (or refer to) the elements of P.

---

[3]We assume that the elements of propositions can include, for example, individuals, properties, relations and other propositions. Other metaphysical assumptions are, of course, possible. For example, it is arguable that the constituents of propositions include *individual concepts* (in the Fregian sense) rather than individuals themselves; and so on. Fortunately, it is not necessary to enter into these abstruse issues to make the points that are relevant to the systematicity problem. All we really need is that propositions have internal structure, and that, characteristically, the internal structure of complex mental representations corresponds, in the appropriate way, to the internal structure of the propositions that they express.

Systematicity

The Classical explanation of systematicity assumes that C holds by nomological necessity; it expresses a *psychological law* that subsumes all systematic minds. It should be fairly clear why systematicity is readily explicable on the assumptions, first, that mental representations satisfy C, and, second, that mental processes have access to the constituent structure of mental representations. Thus, for example, since C implies that anyone who can represent a proposition can, ipso facto, represent its elements, it implies, in particular, that anyone who can represent the proposition that John loves the girl can, ipso facto, represent John, the girl and the two-place relation *loving*. Notice, however, that the proposition that *the girl loves John* is *also* constituted by these same individuals/relations. So, then, assuming that the processes that integrate the mental representations that express propositions have access to their constituents, it follows that anyone who can represent John's loving the girl can also represent the girl's loving John. Similarly, suppose that the constituents of the mental representation that gets tokened when one thinks that P&Q&R and the constituents of the mental representation that gets tokened when one thinks that P&Q both include the mental representation that gets tokened when one thinks that P. And suppose that the mental processes that mediate the drawing of inferences have access to the constituent structure of mental representations. Then it should be no surprise that anyone who can infer P from P&Q&R can likewise infer P from P&Q.

To summarize: the Classical solution to the systematicity problem entails that (i) systems of mental representation satisfy C (a fortiori, complex mental representations have Classical constituents); and (ii) mental processes are sensitive to the constituent structure of mental representations. We can now say quite succinctly what our claim against Smolensky will be: on the one hand, the cognitive architecture he endorses does not provide for mental representations with Classical constituents; on the other hand, he provides no suggestion as to how mental processes could be structure sensitive unless mental representations have Classical constituents; and, on the third hand (as it were) he provides no suggestion as to how minds could be systematic if mental processes aren't structure sensitive. So his reply to Fodor and Pylyshyn fails.

Most of the rest of the paper will be devoted to making this analysis stick.

II. Weak compositionality

Smolensky's views about "weak" compositional structure are largely inexplicit and must be extrapolated from his "coffee story", which he tells in both of the papers under discussion (and also in 1988a). We turn now to considering this story.

Smolensky begins by asking how we are to understand the relation between the mental representation COFFEE and the mental representation CUP WITH COFFEE.[4] His answer to this question has four aspects that are of present interest:

(i) COFFEE and CUP WITH COFFEE are activity vectors (according to Smolensky's weak compositional account, this is true of the mental representations corresponding to all commonsense concepts; whether it also holds for (for example) technical concepts won't matter for what follows). A vector is, of course, a magnitude with a certain direction. A pattern of activity over a group of "units" is a state consisting of the members of the group each having an activation value of 1 or 0.[5] Activity vectors are representations of such patterns of activity.

(ii) CUP WITH COFFEE representations contain COFFEE representations as (non-Classical)[6] constituents in the following sense: they contain them as *component* vectors. By stipulation, **a** is a component vector of **b**, if there is a vector **x** such that **a** + **x** = **b** (where "+" is the operation of vector addition). More generally, according to Smolensky, the relation between vectors and their non-Classical constituents is that the former are derivable from the latter by operations of vector analysis.

(iii) COFFEE representations and CUP WITH COFFEE representations are activity vectors over units which represent microfeatures (units like BROWN, LIQUID, MADE OF PORCELAIN, etc.).

---

[4]The following notational conventions will facilitate the discussion: we will follow standard practice and use capitalized English words and sentences as canonical names for mental representations. (Smolensky uses italicized English expressions instead.) We stipulate that the semantic value of a mental representation so named is the semantic value of the corresponding English word or sentence, and we will italicize words or sentences that denote semantic values. So, for example, COFFEE is a mental representation that expresses (the property of being) *coffee* (as does the English word "coffee"); JOHN LOVES THE GIRL is a mental representation that expresses the proposition that *John loves the girl*; and so forth. It is important to notice that our notation allows that the mental representation JOHN LOVES THE GIRL can be atomic and the mental representation COFFEE can be a complex symbol. That is, capitalized expressions should be read as the names of mental representations rather than as structural descriptions.

[5]Smolensky apparently allows that units may have continuous levels of activation from 0 to 1. In telling the coffee story, however, he generally assumes bivalence for ease of exposition.

[6]As we shall see below, when an activity vector is tokened, its component vectors typically are not. So the constituents of a complex vector are, ipso facto, non-Classical.

(iv) COFFEE (and, presumably, any other representation vector) is *context dependent*. In particular, the activity vector that is the COFFEE representation in CUP WITH COFFEE *is distinct from* the activity vector that is the COFFEE representation in, as it might be, GLASS WITH COFFEE or CAN WITH COFFEE. Presumably this means that the vector in question, with no context specified, does not give necessary conditions for being *coffee*. (We shall see later that Smolensky apparently holds that it doesn't specify sufficient conditions for being *coffee* either).

Claims i and ii introduce the ideas that mental representations are activity vectors and that they have (non-Classical) constituents. These ideas are neutral with respect to the distinction between strong and weak compositionality so we propose to postpone discussing them until section III. Claim iii, is, in our view, a red herring. The idea that there are microfeatures is orthogonal both to the question of systematicity and to the issues about compositionality. We therefore propose to discuss it only very briefly. It is claim iv that distinguishes the strong from the weak notion of compositional structure: a representation has weak compositional structure iff it contains context-dependent constituents. We propose to take up the question of context dependent-representation here.

We commence by reciting the coffee story (in a slightly condensed form).

Since, following Smolensky, we are assuming heuristically that units have bivalent activity levels, vectors can be represented by ordered sets of zeros (indicating that a unit is "off") and ones (indicating that a unit is "on"). Thus, Smolensky says, the CUP WITH COFFEE representation might be the following activity vector over microfeatures:

1–UPRIGHT CONTAINER
1–HOT LIQUID
0–GLASS CONTACTING WOOD[7]
1–PORCELAIN CURVED SURFACE
1–BURNT ODOR

[7]Notice that this microfeature is "off" in CUP WITH COFFEE, so it might be wondered why Smolensky mentions it at all. The explanation may be this: operations of vector combination apply only to vectors of the same dimensionality. In the context of the weak constituency story, this means that you can only combine vectors that are activity patterns *over the same units*. It follows that a component vector must contain the same units (though, possibly at different levels of activation) as the vectors with which it combines. Thus if GRANNY combines with COFFEE to yield GRANNY'S COFFEE, GRANNY must contain activation levels for all the units in COFFEE and vice versa. In the present example, it may be that CUP WITH COFFEE is required to contain a 0-activation level for GLASS CONTACTING WOOD to accommodate cases where it is a component of some other vector. Similarly with OBLONG SILVER OBJECT (below) since cups with coffee often have spoons in them.

1–BROWN LIQUID CONTACTING PORCELAIN
1–PORCELAIN CURVED SURFACE
0–OBLONG SILVER OBJECT
1–FINGER-SIZED HANDLE
1–BROWN LIQUID WITH CURVED SIDES AND BOTTOM[8]

This vector, according to Smolensky, contains a COFFEE representation as a constituent. This constituent can, he claims, be derived from CUP WITH COFFEE by subtracting CUP WITHOUT COFFEE from CUP WITH COFFEE. The vector that is the remainder of this subtraction will be COFFEE.

The reader will object that this treatment presupposes that CUP WITHOUT COFFEE is a constituent of CUP WITH COFFEE. Quite so. Smolensky is explicit in claiming that "the pattern or vector representing *cup with coffee* is composed of a vector that can be identified as a particular distributed representation of *cup without coffee* with a representation with the content *coffee*" (1988b: p. 10).

One is inclined to think that this must surely be wrong. If you combine a representation with the content *cup without coffee* with a representation with the content *coffee*, you get not a representation with the content *cup with coffee* but rather a representation with the self-contradictory content *cup without coffee with coffee*. Smolensky's subtraction procedure appears to confuse the representation of *cup without coffee* (viz. CUP WITHOUT COFFEE) with the representation of *cup* without the representation of *coffee* (viz. CUP). CUP WITHOUT COFFEE expresses the content *cup without coffee*; CUP combines consistently with COFFEE. But nothing does both.

On the other hand, it must be remembered that Smolensky's mental representations are advertised as context dependent, hence non-compositional. Indeed, we are given *no clue at all* about what sorts of relations between the semantic properties of complex symbols and the semantic properties of their constituents his theory acknowledges. Perhaps in a semantics where constituents don't contribute their contents to the symbols they belong to, it's all right after all if CUP WITH COFFEE has CUP WITHOUT COFFEE (or, for that matter, PRIME NUMBER, or GRANDMOTHER, or FLYING SAUCER or THE LAST OF THE MOHICANS) among its constituents.

---

[8]Presumably Smolensky does not take this list to be exhaustive, but we don't know how to continue it. Beyond the remark that although the microfeatures in his examples correspond to "... nearly sensory-level representation[s] ..." that is "not essential", Smolensky provides no account at all of what determines which contents are expressed by microfeatures. The question thus arises why Smolensky assumes that COFFEE is not itself a microfeature. In any event, Smolensky repeatedly warns the reader not to take his examples of microfeatures very seriously, and we don't.

In any event, to complete the story, Smolensky gives the following features for CUP WITHOUT COFFEE:

1–UPRIGHT CONTAINER
0–HOT LIQUID
0–GLASS CONTACTING WOOD
1–PORCELAIN CURVED SURFACE
0–BURNT ODOR
0–BROWN LIQUID CONTACTING PORCELAIN
1–PORCELAIN CURVED SURFACE
0–OBLONG SILVER OBJECT
1–FINGER-SIZED HANDLE
0–BROWN LIQUID WITH CURVED SIDES AND BOTTOM etc.

Subtracting this vector from CUP WITH COFFEE, we get the following COFFEE representation:

0–UPRIGHT CONTAINER
1–HOT LIQUID
0–GLASS CONTACTING WOOD
0–PORCELAIN CURVED SURFACE
1–BURNT ODOR
1–BROWN LIQUID CONTACTING PORCELAIN
0–PORCELAIN CURVED SURFACE
0–OBLONG SILVER OBJECT
0–FINGER-SIZED HANDLE
1–BROWN LIQUID WITH CURVED SIDES AND BOTTOM

That, then, is Smolensky's "coffee story".

Comments

(i) Microfeatures

It's common ground in this discussion that the explanation of systematicity must somehow appeal to relations between complex mental representations and their constituents (on Smolensky's view, to combinatorial relations among vectors). The issue about whether there are microfeatures is entirely orthogonal; it concerns only the question *which properties the activation states of individual units express*. (To put it in more Classical terms, it concerns the question which symbols constitute the *primitive vocabulary* of the system of mental representations.) If there are microfeatures, then the activation states of individual units are constrained to express only (as it might be) "sensory" properties (1987: p. 146). If there aren't, then activation states of individual units can express not only such properties as *being brown* and *being hot*, but

also such properties as *being coffee*. It should be evident upon even casual reflection that, whichever way this issue is settled, the constituency question—viz., the question how the representation COFFEE relates to the representation CUP WITH COFFEE—remains wide open. We therefore propose to drop the discussion of microfeatures in what follows.

(iv) Context-dependent representation

As far as we can tell, Smolensky holds that the representation of *coffee* that he derives by subtraction from CUP WITH COFFEE is context dependent in the sense that it need bear no more than a "family resemblance" to the vector that represents *coffee* in CAN WITH COFFEE, GLASS WITH COFFEE, etc. There is thus no single vector that counts as *the* COFFEE representation, hence no single vector that is a component of all the representations which, in a Classical system would have COFFEE as a Classical constituent.

Smolensky himself apparently agrees that this is the wrong sort of constituency to account for systematicity and related phenomena. As he remarks, "a true constituent can move around and fill any of a number of different roles in different structures" (1988b: p. 11) and the connection between constituency and systematicity would appear to turn on this. For example, the solution to the systematicity problem mooted in section I depends exactly on the assumption that tokens of the representation type JOHN express the same content in the context LOVES THE GIRL that they do in the context THE GIRL LOVES; (viz., that they pick out *John*, who is an element both of the proposition *John loves the girl* and of the proposition *the girl loves John*.) It thus appears, prima facie, that the explanation of systematicity requires context-independent constituents.

How, then, does Smolensky suppose that the assumption that mental representations have weak compositional structure, that is that mental representation is context dependent, bears on the explanation of systematicity? He simply doesn't say. And we don't have a clue. In fact, having introduced the notion of weak compositional structure, Smolensky to all intents and purposes drops it in favor of the notion of strong compositional structure, and the discussion of systematicity is carried out entirely in terms of the latter. What, then, he takes the relation between weak and strong compositional structure to be,—and, for that matter, which kind of structure he actually thinks that mental representations have[9]—is thoroughly unclear.

---

[9]They can't have both; either the content of a representation is context dependent or it's not. So, if Smolensky does think that you need strong compositional structure to explain systematicity, and that weak compositional structure is the kind that Connectionist representations have, then it would seem that he *thereby* grants Fodor and Pylyshyn's claim that Connectionist representations can't explain systematicity. We find this all very mysterious.

In fact, quite independent of its bearing on systematicity, the notion of weak compositional structure as Smolensky presents it is of very dubious coherence. We close this section with a remark or two about this point.

It looks as though Smolensky holds that the COFFEE vector that you get by subtraction from CUP WITH COFFEE is not a COFFEE representation when it stands alone. "This representation is indeed a representation of coffee, but [only?] in a very particular context: the context provided by *cup* [i.e. CUP]" (1987: p. 147). If this is the view, it has bizarre consequences. Take a liquid that has the properties specified by the microfeatures that comprise COFFEE in isolation, but that isn't coffee. Pour it into a cup, et voila! it *becomes* coffee by semantical magic.

Smolensky explicitly doesn't think that the vector COFFEE that you get from CUP WITH COFFEE gives necessary conditions for being coffee, since you'd get a different COFFEE vector by subtraction from, say, GLASS WITH COFFEE. And the passage just quoted suggests that he thinks it doesn't give sufficient conditions either. But, then, if the microfeatures associated with COFFEE are neither necessary nor sufficient for being *coffee*[10] the question arises what, according to this story, *does* makes a vector a COFFEE representation; when does a vector have the content *coffee*?

As far as we can tell, Smolensky holds that what makes the COFFEE component of CUP WITH COFFEE a representation with the content *coffee* is that it is distributed over units representing certain microfeatures *and* that it figures as a component vector of a vector which is a CUP WITH COFFEE representation. As remarked above, we are given no details at all about this reverse compositionality according to which the embedding vector determines the contents of its constituents; how it is supposed to work isn't even discussed in Smolensky's papers. But, in any event, a regress threatens since the question now arises: if being a component of a CUP OF COFFEE representation is required to make a vector a *coffee* representation, what is required to make a vector a *cup of coffee* representation? Well, presumably CUP OF COFFEE represents *cup of coffee* because it involves the microfeatures it does *and* because it is a component of still another vector; perhaps one that is a THERE IS A CUP OF COFFEE ON THE TABLE representation. Does this go on forever? If it doesn't, then presumably there are some vectors which aren't constituents of any others. But now, what determines *their* contents? Not the contents of their constituents because, by assumption, Smolensky's semantics isn't compositional (CUP WITHOUT COFFEE is a constituent of CUP WITH COFFEE, etc.). And not the vectors that they

[10] If they were necessary and sufficient, COFFEE wouldn't be context dependent.

are constituents of, because, by assumption, there aren't any of those.

We think it is unclear whether Smolensky has a coherent story about how a system of representations could have weak compositional structure.

What, in light of all this, leads Smolensky to embrace his account of weak compositionality? Here's one suggestion: perhaps Smolensky confuses being a representation of a cup with coffee with being a CUP WITH COFFEE representation. Espying some cup with coffee on a particular occasion, in a particular context, one might come to be in a mental state that represents it as having roughly the microfeatures that Smolensky lists. That mental state would then be a representation of a cup with coffee in this sense: there is a cup of coffee that it's a mental representation of. But it wouldn't, of course, follow, that it's a CUP WITH COFFEE representation; and the mental representation of that cup with coffee might be quite different from the mental representation of the cup with coffee that you espied on some other occasion or in some other context. So *which mental representation a cup of coffee gets is context dependent*, just as Smolensky says. But that doesn't give Smolensky what he needs to make mental representations themselves context dependent. In particular, from the fact that cups with coffee get different representations in different contexts, it patently doesn't follow that the mental symbol that represents something as *being* a cup of coffee in one context might represent something as being something else (a giraffe say, or The Last of The Mohicans) in some other context. We doubt that anything will give Smolensky that, since we know of no reason to suppose that it is true.

In short, it is natural to confuse the true but uninteresting thought that how you mentally represent some coffee depends on the context, with the much more tendentious thought that the mental representation COFFEE is context dependent. Assuming that he is a victim of this confusion makes sense of many of the puzzling things that Smolensky says in the coffee story. Notice, for example, that all the microfeatures in his examples express more or less perceptual properties (cf. Smolensky's own remark that his microfeatures yield a "nearly sensory level representation"). Notice, too, the peculiarity that the microfeature "porcelain curved surface" occurs *twice* in the vector for CUP WITH COFFEE, COFFEE, CUP WITHOUT COFFEE and the like. Presumably, what Smolensky has in mind is that, when you look at a cup, you get to see two curved surfaces, one going off to the left and the other going off to the right.

Though we suspect this really is what's going on, we won't pursue this interpretation further since, if it's correct, then the coffee story is completely irrelevant to the question of what kind of constituency relation a COFFEE representation bears to a CUP WITH COFFEE; and that, remember, is the question that bears on the issues about systematicity.

III. Strong compositional structure

So much, then, for "weak" compositional structure. Let us turn to Smolensky's account of "strong" compositional structure. Smolensky says that:

> A true constituent can move around and fill any of a number of different roles in different structures. Can *this* be done with vectors encoding distributed representations, and be done in a way that doesn't amount to simply implementing symbolic syntactic constituency? The purpose of this section is to describe research showing that the answer is affirmative. (1988b: p. 11)

The idea that mental representations are activity vectors over units, and the idea that some mental representations have other mental representations as components, is common to the treatment of both weak and strong compositional structure. However, Smolensky's discussion of the latter differs in several respects from his discussion of the former. First, units are explicitly supposed to have continuous activation levels between 0 and 1; second, he does not invoke the idea of microfeatures when discussing strong compositional structure; third, he introduces a new vector operation (multiplication) to the two previously mentioned (addition and subtraction); fourth, and most important, strong compositional structure does not invoke—indeed, would appear to be incompatible with—the notion that mental representations are context dependent. So strong compositional structure does not exhibit the incoherences of Smolensky's theory of context-dependent representation.

We will proceed as follows. First we briefly present the notion of strong compositional structure. Then we shall turn to criticism.

Smolensky explains the notion of strong compositional structure, in part, by appeal to the ideas of a tensor product representation and a superposition representation. To illustrate these ideas, consider how a connectionist machine might represent four-letter English words. Words can be decomposed into roles (viz., ordinal positions that letters can occupy) and things that can fill these roles (viz., letters). Correspondingly, the machine might contain activity vectors over units which represent the relevant roles (i.e., over the *role units*) and activity vectors over units which represent the fillers (i.e., over the *filler units*). Finally, it might contain activity vectors over units which represent *filled roles* (i.e., letters in letter positions); these are the *binding units*. The key idea is that the activity vectors over the binding units might be tensor products of activity vectors over the role units and the filler units. The representation of a word would then be a superposition vector over the binding units; that is, a vector that is arrived at by superimposing the tensor product vectors.

The two operations used here to derive complex vectors from component vectors are vector multiplication in the case of tensor product vectors and vector addition in the case of superposition vectors. These are iterative operations in the sense that activity vectors that result from the multiplication of role vectors and filler vectors might themselves represent the fillers of roles in more complex structures. Thus, a tensor product which represents the word "John" as *"J" in first position, "o" in second position ... etc.* might itself be bound to the representation of a syntactical function to indicate, for example, that "John" has the role subject-of in "John loves the girl". Such tensor product representations could themselves be superimposed over yet another group of binding units to yield a superposition vector which represents the bracketing tree (John) (loves (the girl)).

It is, in fact, unclear whether this sort of apparatus is adequate to represent all the semantically relevant syntactic relations that Classical theories express by using bracketing trees with Classical constituents. (There are, for example, problems about long-distance binding relations, as between quantifiers and bound variables.) But we do not wish to press this point. For present polemical purposes, we propose simply to assume that each Classical bracketing tree can be coded into a complex vector in such fashion that the constituents of the tree correspond in some regular way to components of the vector.

But this is not, of course, to grant that either tensor product or superposition vectors *have* Classical constituent structure. In particular, from the assumptions that bracketing trees have Classical constituents and that bracketing trees can be coded by activity vectors, it does *not* follow that activity vectors have Classical constituents. On the contrary, a point about which Smolensky is himself explicit is vital in this regard: the components of a complex vector need not even correspond to patterns of activity over units actually in the machine. As Smolensky puts it, the activity states of the filler and role units can be "imaginary" even though the ultimate activity vectors—the ones which do not themselves serve as filler or role components of more complex structures—must be actual activity patterns over units in the machine. Consider again our machine for representing four-letter words. The superposition pattern that represents, say, the word "John" will be an activity vector actually realized in the machine. However, the activity vector representing "J" will be merely imaginary, as will the activity vector representing *the first letter position*. Similarly for the tensor product activity vector representing *"J" in the first letter position*. The only pattern of activity that will be *actually tokened* in the machine is the superposition vector representing "John".

These considerations are of central importance for the following reason. Smolensky's main strategy is, in effect, to invite us to consider the compo-

nents of tensor product and superposition vectors to be analogous to the Classical constituents of a complex symbol; hence to view them as providing a means by which connectionist architectures can capture the causal and semantic consequences of Classical constituency in mental representations. However, the components of tensor product and superposition vectors differ from Classical constituents in the following way: when a complex Classical symbol is tokened, its constituents are tokened. When a tensor product vector or superposition vector is tokened, its components are not (except per accidens). The implication of this difference, from the point of view of the theory of mental processes, is that whereas the Classical constituents of a complex symbol are, ipso facto, available to contribute to the causal consequences of its tokenings—in particular, they are available to provide domains for mental processes—the components of tensor product and superposition vectors can have no causal status as such. What is merely imaginary can't make things happen, to put this point in a nutshell.

We will return presently to what all this implies for the treatment of the systematicity problem. There is, however, a preliminary issue that needs to be discussed.

We have seen that the components of tensor product/superposition vectors, unlike Classical constituents, are not, in general, tokened whenever the activity vector of which they are the components is tokened. It is worth emphasizing, in addition, the familiar point that there is, in general, no *unique* decomposition of a tensor product or superposition vector into components. Indeed, given that units are assumed to have continuous levels of activation, there will be *infinitely* many decompositions of a given activity vector. One might wonder, therefore, what sense there is in talk of *the* decomposition of a mental representation into significant constituents given the notion of constituency that Smolensky's theory provides.[11]

Smolensky replies to this point as follows. Cognitive systems will be dynamical systems; there will be dynamic equations over the activation values of individual units, and these will determine certain regularities over activity vectors. Given the dynamical equations of the system, certain decompositions can be especially useful for "explaining and understanding" its behavior. In this sense, the dynamics of a system may determine "normal modes" of decomposition into components. So, for example, though a given superposition vector can, in principle, be taken to be the sum of many different sets of vectors, yet it may turn out that we get a small group of sets—even a unique

[11]The function of the brackets in a Classical bracketing tree is precisely to exhibit its decomposition into constituents; and when the tree is well formed this decomposition will be unique. Thus, the bracketing of "(John) (loves) (the girl)" implies, for example, both that "the girl" is a constituent and that "loves the" is not.

set—when we decompose in the direction of normal modes; and likewise for decomposing tensor product vectors. The long and short is that *it could, in principle, turn out* that, given the (thus far undefined) normal modes of a dynamical cognitive system, complex superposition vectors will have it in common with Classical complex symbols that they have a unique decomposition into semantically significant parts. Of course, it also could turn out that they don't, and no ground for optimism on this point has thus far been supplied.

Having noted this problem, however, we propose simply to ignore it. So here is where we now stand: by assumption (though quite possibly contrary to fact), tensor product vectors and superposition vectors can code constituent structure in a way that makes them adequate vehicles for the expression of propositional content; and, by assumption (though again quite possibly contrary to fact), the superposition vectors that cognitive theories acknowledge have a unique decomposition into semantically interpretable tensor product vectors which, in turn, have a unique decomposition into semantically interpretable filler vectors and role vectors; so it's determinate which proposition a given complex activity vector represents.

Now, assuming all this, what about the systematicity problem?

The first point to make is this: if tensor product/superposition vector representation solves the systematicity problem, the solution must be quite different from the Classical proposal sketched in section I. True tensor product vectors and superposition vectors "have constituents" in some suitably extended sense: tensor product vectors have semantically evaluable components, and superposition vectors are decomposable into semantically evaluable tensor product vectors. But the Classical solution to the systematicity problem assumes that *the constituents of mental representations have causal roles*; that they provide domains for mental processes. The Classical constituents of a complex symbol thus contribute to determining the causal consequences of the tokening of that symbol, and it seems clear that the "extended" constituents of a tensor product/superposition representation can't do that. On the contrary, the components of a complex vector are typically not even tokened when the complex vector itself is tokened; they are simply constituents into which the complex vector *could be* resolved consonant with decomposition in the direction of normal modes. But, to put it crudely, the fact that six *could be* represented as "3 × 2" cannot, in and of itself, affect the causal processes in a computer (or a brain) in which six *is* represented as "6". Merely counterfactual representations have no causal consequences; only actually tokened representations do.

Smolensky is, of course, sensitive to the question whether activity vectors really do have constituent structure. He defends at length the claim that he

has not contorted the notion of constituency in claiming that they do. Part of this defense adverts to the role that tensor products and superpositions play in physical theory:

> The state of the atom, like the states of all systems in quantum theory, is represented by a vector in an abstract vector space. Each electron has an internal state (its "spin"); it also has a role it plays in the atom as a whole: it occupies some "orbital", essentially a cloud of probability for finding it at particular places in the atom. The internal state of an electron is represented by a "spin vector"; the orbital or role of the electron (part) in the atom (whole) is represented by another vector, which describes the probability cloud. The vector representing the electron as situated in the atom is the tensor product of the vector representing the internal state of the electron and the vector representing its orbital. The atom as a whole is represented by a vector that is the sum or superposition of vectors, each of which represents a particular electron in its orbital ... (1988b: pp. 19–20)

"So," Smolensky adds, "someone who claims that the tensor product representational scheme distorts the notion of constituency has some explaining to do" (1988b: p. 20).

The physics lesson is greatly appreciated; but it is important to be clear on just what it is supposed to show. It's not, at least for present purposes, in doubt that tensor products *can represent* constituent structure. The relevant question is whether tensor product representations *have* constituent structure; or, since we have agreed that they may be said to have constituent structure "in an extended sense", it's whether they have the kind of constituent structure to which causal processes can be sensitive, hence the kind of constituent structure to which an explanation of systematicity might appeal.[12] But we have already seen the answer to *this* question: the constituents of complex activity vectors typically aren't "there", so if the causal consequences of tokening a complex vector are sensitive to its constituent structure, that's a miracle.

We conclude that assuming that mental representations are activation vectors does not allow Smolensky to endorse the Classical solution of the systematicity problem. And, indeed, we think Smolensky would grant this since he admits up front that mental processes will not be causally sensitive to the strong compositional structure of mental representations. That is, he acknowledges that the constituents of complex mental representations play no causal

[12]It's a difference between psychology and physics that whereas psychology is about the casual laws that govern tokenings *of (mental) representations*, physics is about the causal laws that govern (not mental representations but) atoms, electrons and the like. Since *being a representation* isn't a property in the domain of physical theory, the question whether mental representations have constituent structure has no analog in physics.

role in determining what happens when the representations get tokened. "... Causal efficacy was not my goal in developing the tensor product representation ..." (1988b: p. 21). What are causally efficacious according to connectionists are the activation values of individual units; the dynamical equations that govern the evolution of the system will be defined over these. It would thus appear that Smolensky must have some *non*-Classical solution to the systematicity problem up his sleeve; some solution that does *not* depend on assuming mental processes that are causally sensitive to constituent structure. So then, after all this, what *is* Smolensky's solution to the systematicity problem?

Remarkably enough, *Smolensky doesn't say*. All he does say is that he "hypothesizes ... that ... the systematic effects observed in the processing of mental representations arise because the evolution of vectors can be (at least partially and approximately) explained in terms of the evolution of their components, even though the precise dynamical equations apply [only] to the individual numbers comprising the vectors and [not] at the level of [their] constituents—i.e. even though the constituents are not causally efficacious" (1988b: p. 21).

It is left unclear how the constituents ("components") of complex vectors are to explain their evolution (even partially and approximately) when they are, by assumption, at best causally inert and, at worst, merely imaginary. In any event, what Smolensky clearly does think is causally responsible for the "evolution of vectors" (and hence for the systematicity of cognition) are unspecified processes that affect the states of activation of the individual units (the neuron analogs) out of which the vectors are composed. So, then, as far as we can tell, the proposed connectionist explanation of systematicity (and related features of cognition) comes down to this: Smolensky "hypothesizes" that systematicity is somehow a consequence of underlying neural processes.[13] Needless to say, if that *is* Smolensky's theory, it is, on the one hand, certainly true, and, on the other hand, not intimately dependent upon his long story about fillers, binders, tensor products, superposition vectors and the rest.

By way of rounding out the argument, we want to reply to a question raised by an anonymous *Cognition* reviewer, who asks: "... couldn't Smolensky easily build in mechanisms to accomplish the matrix algebra oper-

[13]More precisely: we take Smolensky to be claiming that there is some property D, such that if a dynamical system has D its behavior is systematic, and such that human behavior (for example) is caused by a dynamical system that has D. The trouble is that this is a platitude since it is untendentious that human behavior is systematic, that its causation by the nervous system is lawful, and that the nervous system is dynamical. The least that has to happen if we are to have a substantive connectionist account of systematicity is: first, it must be made clear what property D is, and second it must be shown that D is a property that connectionist systems can have by law. Smolensky's theory does nothing to meet either of these requirements.

ations that would make the necessary vector explicit (or better yet, from his point of view, ... mechanisms that are sensitive to the imaginary components without literally making them explicit in some string of units)?"[14] But this misses the point of the problem that systematicity poses for connectionists, which is not to show that systematic cognitive capacities are *possible* given the assumptions of a connectionist architecture, but to explain how systematicity could be *necessary*—how it could be a *law* that cognitive capacities are systematic--given those assumptions.[15]

No doubt it is possible for Smolensky to wire a network so that it supports a vector that represents aRb if and only if it supports a vector that represents bRa; and perhaps it is possible for him to do that without making the imaginary units explicit[16] (though there is, so far, no proposal about how to ensure this for *arbitrary* a, R and b). The trouble is that, although the architecture permits this, it equally permits Smolensky to wire a network so that it supports a vector that represents aRb if and only if it supports a vector that represents zSq; or, for that matter, if and only if it supports a vector that represents The Last of The Mohicans. The architecture would appear to be absolutely indifferent as among these options.

Whereas, as we keep saying, in the Classical architecture, if you meet the conditions for being able to represent aRb, *YOU CANNOT BUT MEET THE CONDITIONS FOR BEING ABLE TO REPRESENT bRa*; the architecture won't let you do so because (i) the representation of a, R and b

[14]Actually, Smolensky is forced to choose the second option. To choose the first would, in effect, be to endorse the Classical requirement that tokening a symbol implies tokening its constituents; in which case, the question arises once again why such a network isn't an implementation of a language of thought machine. Just as Smolensky mustn't allow the representations of roles, fillers and binding units to be subvectors of superposition vectors if he is to avoid the "implementation" horn of the Fodor/Pylyshyn dilemma, so too he must avoid postulating mechanisms that make role, filler and binding units explicit (specifically, accessible to mental operations) whenever the superposition vectors are tokened. Otherwise he again has symbols with Classical constituents and raises the question why the proposed device isn't a language of thought machine. Smolensky's problem is that the very feature of his representations that make them wrong for explaining systematicity (viz., that their constituents are allowed to be imaginary) is the one that they have to have to assure that they aren't Classical.

[15]Fodor and Pylyshyn were very explicit about this. See, for example, 1988: p. 48.

[16]Terence Horgan remarks (personal communication) "... often there are two mathematically equivalent ways to calculate the time-evolution of a dynamical system. One is to apply the relevant equations directly to the numbers that are elements of a single total vector describing the initial state of the system. Another way is to mathematically decompose that vector into component normal-mode vector, then compute the time-evolution of each [of these] ... and then take the later state of the system to be described by a vector that is the superposition of the resulting normal-mode vectors." Computations of the former sort are supposed to be the model for operations that are "sensitive" to the components of a mental representation vector without recovering them. (Even in the second case, it's the theorist who recovers them in the course of the computations by which he makes his predictions. This does not, of course, imply that the constituents thus "recovered" participate in causal processes in the system under analysis.)

are constituents of the representation of aRb, and (ii) you have to token the constituents of the representations that you token, so Classical constituents can't be just imaginary. So then: it is *built into* the Classical picture that you can't think aRb unless you are able to think bRa, but the Connectionist picture is *neutral* on whether you can think aRb even if you can't think bRa. But it is a law of nature that you can't think aRb if you can't think bRa. So the Classical picture explains systematicity and the Connectionist picture doesn't. So the Classical picture wins.

Conclusion

At one point in his discussion, Smolensky makes some remarks that we find quite revealing: he says that, even in cases that are paradigms of Classical architectures (LISP machines and the like), "... we normally think of the 'real' causes as physical and far below the symbolic level ..." Hence, even in Classical machines, the sense in which operations at the symbol level are real causes is just that "... there is ... a complete and precise algorithmic (temporal) story to tell about the states of the machine described ..." at that level (1988b: p. 20). Smolensky, of course, denies that there is a "... comparable story at the symbolic level in the human cognitive architecture ... that is a difference with the Classical view that I have made much of. *It may be that a good way to characterize the difference is in terms of whether the constituents in mental structure are causally efficacious in mental processing*" (1988b: p. 20; our emphasis).

We say that this is revealing because it suggests a diagnosis: it would seem that Smolensky has succumbed to a sort of generalized epiphenomenalism. The idea is that even Classical constituents participate in causal processes solely by virtue of their physical microstructure, so even on the Classical story it's what happens at the neural level that *really* counts. Though the evolution of vectors can perhaps be explained in a predictively adequate sort of way by appeal to macroprocesses like operations on constituents, still if you want to know what's *really* going on—if you want the *causal* explanation—you need to go down to the "precise dynamical equations" that apply to activation states of units. That intentional generalizations can only approximate these precise dynamical equations is among Smolensky's recurrent themes. By conflating the issue about "precision" with the issue about causal efficacy, Smolensky makes it seem that to the extent that macrolevel generalizations are imprecise, to that extent macrolevel processes are epiphenomenal.

It would need a philosophy lesson to say all of what's wrong with this. Suffice it for present purposes that the argument iterates in a way that

Smolensky ought to find embarrassing. No doubt, we do get greater precision when we go from generalizations about operations on constituents to generalizations about operations on units. But if that shows that symbol-level processes aren't really causal, then it must be that unit-level processes aren't really causal either. After all, we get *still more* precision when we go down from unit-sensitive operations to molecule-sensitive operations, and more precision yet when we go down from molecule-sensitive operations to quark-sensitive operations. The moral is not, however, that the causal laws of psychology should be stated in terms of the behavior of quarks. Rather, the moral is that whether you have a level of causal explanation is a question, not just of how much precision you are able to achieve, but also of *what generalizations you are able to express*. The price you pay for doing psychology at the level of units is that you lose causal generalizations that symbol-level theories are able to state. Smolensky's problems with capturing the generalizations about systematicity provide a graphic illustration of these truths.

It turns out, at any event, that there is a crucial caveat to Smolensky's repeated claim that connectionist mechanisms can reconstruct everything that's interesting about the notion of constituency. Strictly speaking, he claims only to reconstruct whatever is interesting about constituents *except their causes and effects*. The explanation of systematicity turns on the causal role of the constituents of mental representations and is therefore among the casualties. Hilary Putnam, back in the days when he was still a Metaphysical Realist, used to tell a joke about a physicist who actually managed to build a perpetual motion machine; all except for a part that goes back and forth, back and forth, back and forth, forever. Smolensky's explanation of systematicity has very much the character of this machine.

We conclude that Fodor and Pylyshyn's challenge to connectionists has yet to be met. We still don't have *even a suggestion* of how to account for systematicity within the assumptions of connectionist cognitive architecture.

References

Fodor, J., & Pylyshyn, P. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3–71.

Smolensky, P. (1987). The constituent structure of mental states: A reply to Fodor and Pylyshyn. *Southern Journal of Philosophy, 26*, 137–160.

Smolensky, P. (1988a). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1–23.

Smolensky, P. (1988b). Connectionism, constituency and the language of thought. University of Colorado Technical report; also forthcoming in Loewer, B., & Rey, G. (Eds.), *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.