Mind & Language, ISSN: 0268-1064
Vol. 12. No. 2 June 1997, pp 154–177.

Can Connectionists Explain Systematicity?

ROBERT J. MATTHEWS

Abstract: Classicists and connectionists alike claim to be able to explain systematicity. The proposed classicist explanation, I argue, is little more than a promissory note, one that classicists have no idea how to redeem. Smolensky's (1995) proposed connectionist explanation fares little better: it is not vulnerable to recent classicist objections, but it nonetheless fails, particularly if one requires, as some classicists do, that explanations of systematicity take the form of a 'functional analysis'. Nonetheless, there are, I argue, reasons for cautious optimism about the prospects of a connectionist explanation.

Connectionist enthusiasts claim that connectionist architectures can provide the foundation for a new paradigm of computational theories of cognition, one that will supplant the familiar 'classical', i.e. 'language of thought', conception of cognition as the structure-sensitive processing of syntactically structured representations. Classicists, for their part, dismiss such connectionist claims as unbridled fantasy. To hear classicists tell it, the prospects for a connectionist theory of cognition are in fact exceedingly dim: 'the mind cannot be, in its general structure, a Connectionist network' (Fodor and Pylyshyn, 1988, p. 33), or if it can, then connectionists have yet to demonstrate that it can. Connectionists cannot explain, or at least have not shown that they can explain, any of the fundamental aspects of human cognition, notably, systematicity, productivity and inferential coherence. To those connectionists who insist that they can explain cognition, classicists challenge them to provide the requisite demonstration.

In this paper I examine the classicists' claim that they can, but connectionists cannot, explain systematicity. I begin by reviewing the broad outlines of the classicist challenge to connectionists. I then turn to the explanation of systematicity proposed by classicists, arguing that their proposed explanation is in fact nothing more than a promissory note, one that at this point classicists have little idea how to redeem. Next I consider a recent proposal by Smolensky (1995) for a connectionist explanation of system-

I wish to thank Walter Dean, Frances Egan, Brian McLaughlin, Mark Moyer and Jonathan Weinberg for helpful discussions of the issues addressed in this paper.

Address for correspondence: Department of Philosophy, Rutgers University, New Brunswick, NJ 08903, USA.
Email: rjm@rci.rutgers.edu.

aticity. Smolensky's proposal, I argue, is not vulnerable to the objections raised against it by classicists, notably by Fodor and McLaughlin (1990) and McLaughlin (1993a, 1993b); but neither, I argue, is it clear that Smolensky's proposal can in fact provide the needed explanation of systematicity, particularly if one requires, as McLaughlin and apparently many others do, that any adequate explanation must take the form of what Cummins (1983) termed a 'functional analysis'. I conclude with some cautiously optimistic remarks about the prospects for a connectionist explanation of systematicity that does not take such a form.

1. The Challenge to Connectionists[1]

In Fodor and Pylyshyn (1988), the challenge to connectionists was to 'show that the processes which operate on the *representational states* of an organism are those which are specified by a Connectionist architecture' (p. 10). For Fodor and Pylyshyn, this challenge was one that connectionists simply could not meet. Their argument for this conclusion focused on a putative property of cognitive capacities, namely, systematicity. Connectionists, they argued, cannot explain systematicity because they lack a crucial explanatory resource found only in classical architectures, namely, representational states with constituent structure. Fodor and Pylyshyn apparently take the deficit here to be more than explanatory, because they conclude from the fact that connectionist representations lack constituent structure that 'the architecture of the mind is not a Connectionist network' (p. 40). Connectionist networks, they apparently believe, are unable even to exhibit systematicity.

In Fodor and McLaughlin (1990) the challenge to connectionists is put this way: 'to explain the existence of systematic relations among cognitive capacities without assuming that cognitive processes are causally sensitive to the constituent structure of mental representations' (pp. 183–4). Fodor and McLaughlin see this challenge as implying 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' (p. 184). The complaint is no longer that connectionists cannot explain systematicity (except by implementing a classical architecture); rather it is that they have not shown us that they can. The challenge to connectionists is to provide the requisite demonstration.[2]

---

[1] Much of this section is drawn from Matthews, 1991.

[2] The shift towards a burden-of-proof argument evident in Fodor and McLaughlin, 1990, becomes explicit when McLaughlin, 1993a, p. 174, claims that 'connectionists who accept that an adequate theory of cognition must explain systematicity have the burden of proof'.

In mounting their challenge to connectionists, Fodor and Pylyshyn (1988) offer very little by way of a general characterization of systematicity, the explanandum that connectionists are challenged to explain. Instead they offer numerous examples of putatively systematic, i.e. systematically related, cognitive capacities. They point out, for example, that the ability to produce/understand some sentences is related systematically to the ability to produce/understand certain others: you don't find subjects who know how to say in English that John loves the girl but don't know how to say in English that the girl loves John.

Given the way that Fodor and Pylyshyn choose to frame the classicist/connectionist debate, i.e. as a debate regarding cognitive computational architecture, one might reasonably have assumed that systematicity could (and indeed should) be characterized in purely computational terms. But in Fodor and McLaughlin (1990), and especially in McLaughlin (1993a), it becomes evident that systematicity, as these classicists (Fodor et al., as I shall call them) construe it, is *not* amenable to such a characterization. As they construe it, systematicity is a capacity for *intentional*, more specifically propositional attitude, states whose contents are interrelated in certain systematic ways. Thus, McLaughlin (1993a) describes as systematic those cognitive capacities that are '(i) capacities to have intentional states in the same intentional mode (e.g., preference, belief, seeing as), and (ii) the intentional states in question have related contents' (p. 168).

Fodor et al.'s intentional construal of systematicity entails that computational architectures, by themselves, cannot explain systematicity, since architectural descriptions of devices, classical or otherwise, do not entail any particular intentional characterization of these devices.[3] On the proposed construal, systematicity is not, strictly speaking, a property of computational architectures, but of computational architectures *under a particular intentional interpretation*. Hence, only when coupled with a computational construal of intentional states can a computational architecture possibly explain systematicity. Fodor et al. claim to have a classical explanation of systematicity (see below) inasmuch as they have a proposed computational construal of intentional states, more specifically of propositional attitudes, for classical architectures, viz., Fodor's language-of-thought hypothesis (cf. Fodor, 1975, 1987). The challenge to connectionists is to show that they can do as well, though in a way that doesn't rely on this same language-of-thought interpretation of intentional states.

Although Fodor et al.'s intentional construal of systematicity precludes a purely computational explanation, it is important to be clear about the focus of the classicist complaint against connectionism. It is not that connectionists have failed to provide a computational construal of intentional states for connectionist architectures, though clearly this is something they have not done. Rather it is that the computational states of connectionist devices that

[3] For discussion, see Matthews, 1994.

would presumably have to receive an intentional interpretation seemingly lack the properties requisite for an adequate explanation of systematicity. Connectionist architectures seemingly lack the resources to support an explanation of why (or at least how it is possible that) cognitive capacities are systematically related.

2. *Computational Capacity: A Red Herring?*

Not surprisingly, the classicist challenge to connectionists has been widely viewed as having something to do with the computational capacities of connectionist architectures. Claims that the mind cannot be a connectionist network seem to be claims to the effect that connectionist architectures are unable to perform certain computational tasks that classical architectures are (presumably) able to perform. Certainly a number of connectionists have construed the challenge in this way. Chalmers (1990, p. 61), for example, describes a connectionist network that, he claims, provides a 'direct counterexample' to the argument of Fodor and McLaughlin (1990) that to support structure-sensitive processing, representations of constituent structure must contain explicit tokens of the constituents:

> If a representation of 'John loves Michael' is not concatenation of tokens of 'John', 'loves', and 'Michael', they argue, then later processing cannot be sensitive to the compositional structure that is represented. The results presented here show that this conclusion is false. In the distributed representations formed by RAAM, there is no such explicit tokening of the original words. . . . Nevertheless, the representations support systematic processing. *Explicit* constituent structure is not needed for systematicity; implicit structure is enough.

In fact, the computational capacity of connectionist networks is surely not at issue. Of course, computational capacity is *potentially* relevant to the challenge. If connectionist networks were unable to compute the functions that systematic cognitive capacities instantiate, except by implementing a classical architecture, then clearly the challenge to connectionists could not be met. However, this is not an issue that Fodor et al. have ever raised. Fodor et al. have nowhere challenged the by now well-known result that connectionist architectures can compute (or at least approximate to any arbitrary degree) any function computable by means of a classical architecture.[4] Nor do they anywhere assert that this computational power can be achieved only by dint of implementing a classical architecture.

Nevertheless, certain things that Fodor et al. say do seem to imply that

[4] See, for example, Hornik et al., 1989.

they believe that connectionist architectures are limited computationally in a way that classical architectures are not. In dismissing Chalmers's claim that his network provides a 'direct counterexample' to Fodor and McLaughlin's argument that connectionist architectures are incapable of structure-sensitive processing (except by implementing a classical architecture), McLaughlin (1993a, p. 178) denies that Chalmers' network has any bearing whatever on the connectionism/classicism debate. It has no bearing on the debate, we are told, because contrary to what Chalmers claims, the network contains neither syntactically structured representations nor syntax-sensitive processes:

> First, . . . Chalmers's representations only represent syntactic structures; they do not have syntactic structure. (Moreover, they represent syntactic structures, I might add, only by stipulation, which is par for the course for both classical and connectionist AI.) Second, since the representations do not themselves have syntactic structure, they cannot participate in syntax-sensitive processes.

Now, this certainly looks like a claim to the effect that connectionist devices, at least of the sort described by Chalmers, cannot do something that classical devices can, though the parenthetical remark seems to suggest that McLaughlin thinks, surprisingly, that classical-AI devices are no better off in this regard. However, as if to forestall the conclusion that he thinks the failure of connectionist networks to explain systematicity is in some way attributable to computational limitations, McLaughlin (1993a, p. 178) goes on to say the following:

> It is perhaps worth noting here that it has never been an issue in the connectionism/classicism debate whether transitions from inputs of networks to outputs of networks can, in a certain sense, *respect* syntactic transitions: A network of hidden units could, for instance, function as an and-gate. And the leading idea of implementation connectionism is, of course, that connectionist processes can implement basic classical algorithms.

Now, it is not at all clear just what McLaughlin takes himself to be conceding here. The suggestion certainly seems to be that while connectionist architectures can only 'respect' certain syntactic operations (or in some cases even implement them), classical architectures can actually execute them. Such a suggestion would seem to deny the formal computational equivalence results mentioned above, unless, perhaps, McLaughlin believes that these results effectively abstract away from an important difference between connectionist and classical representations, one that would underpin the distinction that he appeals to in his criticism of Chalmers (1990) between a representation's actually having a syntactic structure and its merely representing something with a syntactic structure. Such a difference might

also underpin the conviction that connectionists cannot explain systematicity.

In fact, these formal results do abstract away from a feature of the computational devices being compared with respect to computational capacity: they abstract away from the particular representational schemes employed by the devices. Fodor et al. think it essential to any explanation of systematicity that the representations adverted to in the explanation actually *have* (and not merely represent) constituent structure. However, to see why they think this, what this distinction (between having and representing constituent structure) supposedly comes to, and whether they are justified in their assumption (especially given their recognition that connectionist processes can implement basic classical algorithms), we need to consider both Fodor et al.'s proposed classical explanation of systematicity and their objection to Smolensky's proposal to explain systematicity by means of tensor product representations.

3. *The Classical Explanation of Systematicity*

Classicists, such as Fodor et al., assume that (i) cognitive capacities are generally systematic (in the sense, described above), (ii) it is nomologically necessary (and hence counterfactual supporting) that this is so, (iii) there are psychological mechanisms in virtue of whose functioning this is so; and (iv) an adequate explanation will describe these mechanisms and their functioning.[5] The classical explanation of systematicity proposes to satisfy these four assumptions by identifying the general capacity for systematically related intentional states with the possession of a system of mental representation and a set of (types of) structure sensitive mental processes. Crucially, according to Fodor and McLaughlin (1990, pp. 184–88), the system of mental representation possesses the following properties: (1) it has a compositional syntax and semantics; (2) its complex representations have as proper parts constituents in the sense that when a complex representation is physically tokened, its constituents are also physically tokened; and (3) if one of its complex mental representations expresses a proposition P, then that representation's constituents express (or refer to) the elements of P. The set of (types of) mental processes possesses the following property: (4) all its members are causally sensitive to the constituent structure of representations. Moreover, the system of mental representation and the set of mental processes have these properties, classicists further assume, as a matter of nomological necessity, i.e. as a matter of psychological law.

Subtleties aside, the proposed explanation is just Fodor's Language of Thought doctrine, coupled with the concept-based denotational psychosemantics that Fodor has defended in recent years. On this view, being in an

[5] See Fodor and McLaughlin, 1990, p. 185.

intentional state of intentional mode A and propositional content P (or equivalently having an attitude A to a proposition P), e.g. thinking P, is a matter of having a quasi-linguistic (i.e. sentence-like) complex mental representation S, where S expresses the proposition P and the constituents of S express (or refer to) the elements of P. The most primitive of these constituents are said to be concepts, which express (or refer to) the primitive elements of P. The classical explanation of systematicity is said to follow quite directly from these assumed properties (1–4) of the system of mental representation and the set of mental processes defined over these representations.

Given these assumptions, it is fairly clear, we are told, why, for example, the capacity for thinking that John loves the girl should be systematically related to the capacity for thinking that the girl loves John (Fodor and McLaughlin, 1990, p. 188):

> Since [property (3)] 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.

McLaughlin (1993a, p. 171) concedes that the proposed classical explanation of systematicity issues a number of promissory notes, each of which would have to be redeemed before classicists could claim to have explained systematicity. Most notably, McLaughlin points out, classicists owe us (1) a compositional syntax for the language of thought, (2) an adequate psychosemantics for the language and (3) an adequate computational account of the intentional modes. To these must be added the further requirement that classicists provide us with some rationale for assuming that (4) sentences in the language of thought possess constituent structure such that when a complex representation is physically tokened, its constituents are also physically tokened. Finally, classicists owe us a rationale for assuming that (5) propositions that are the contents of mental representations are structured in such fashion that if a mental representation S expresses a proposition P, then that representation's constituents can be taken to express (or refer to) the elements of P. They owe us a rationale for (5), because it is surely more than a little surprising that the proposed classical explanation of systematicity, which is presumably offered as a hypothesis in empirical psychology, should depend on a particular metaphysical doctrine about the nature of propositions.

All this explanatory debt would seem to overwhelm the classicist claim to have an explanation of systematicity! In fact, Fodor and McLaughlin's

proposed classical explanation of systematicity is in even worse shape than McLaughlin's concessions suggest. Even if we accept the proferred promissory notes, their proposed explanation of the systematic relation between thinking that John loves the girl and thinking that the girl loves John simply does not go through. The assumed properties (1–4) entail that anyone who can represent the proposition that John loves the girl can represent the constitutive *elements* of the proposition that the girl loves John, but (and this is the important point) the assumptions do not entail that this individual can represent the proposition itself. Fodor and McLaughlin concede as much when they say, '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' (p. 188). However, plainly, what is claimed to follow from the assumption does not follow from that assumption alone. What is required is some reason to suppose that from the fact that there are processes that can integrate the contituents in question to form a representation of the proposition that John loves the girl, it follows that there are also processes (perhaps the very same processes) that can integrate these same constituents to form a representation of the proposition that the girl loves John. However, nothing in the explanation warrants this supposition, and yet this is the crux of the proposed explanation. To allow classicists simply to stipulate this supposition would be tantamount to allowing them simply to stipulate that they have an explanation.

One might suppose that one gets the required warrant for this supposition from the assumption, which we are granting here for purposes of argument, that the system of mental representation has a compositional syntax and semantics. The thought might be that, like their natural language counterparts, the mental representations that express the propositions in question share the same syntax; hence, whatever structure-building operations that are available for the one representation are available for the other. This line of argument might work for the case at hand, provided of course, that the syntax (and morphology) of the language of thought does in fact assign identical syntactic (and morphological) structures to both representations (an assumption that seems licensed only by the fact that the syntax of English does so). However, what of cases where the representations don't share the same morphosyntactic structure but nonetheless have semantically related contents by virtue of shared constituent concepts, e.g. representations of the proposition that *girls love John* and the proposition that *John loves girls*? What warrants the supposition that subjects that have the structure-building operations for the former must also have, as a matter of nomological necessity, the structure-building operations for the latter? It appears as if the proposed explanation offers nothing more than a bald stipulation, based on the fact that the English counterparts of these representations are morphosyntactically (and semantically) well-formed. Of course, Fodor and McLaughlin may want to argue that the available structure-building operations are licensed by the syntax (and semantics) of the language of thought. However,

then they owe us a specification of the syntax and semantics for the language of thought. And what is more, they owe us an argument for why, as a matter of nomological necessity, the language of thought has this particular syntax and this particular semantics, rather than any other. For failing that, they have nothing that rises to their own standard for an explanation of systematicity.

Fodor and McLaughlin's proposed explanation of systematicity suffers from a further problem: the explanation would seem to apply equally well to many non-cases. Fodor and McLaughlin (1990) tell us that 'the systematicity of cognition consists of, for example, the fact that organisms that can think aRb can think bRa and vice versa' (p. 185n). However, this seems simply false. I can think the thought that x is the sole member of the singleton set {x}, but I am quite certain that I cannot think the thought that the singleton set {x} is the sole member of x. I have no idea what proposition, if any, the sentence the *singleton set {x} is the sole member of x* expresses. And yet Fodor and McLaughlin's proposed explanation offers a ready explanation of my capacity both to think the thought that I can think and to think the one that I cannot think: I am able to represent the constituent elements of the proposition that I can think, and hence, according to their explanation, I am able to represent the proposition that in fact I cannot think. There is, so far as I can see, only one response available to Fodor and McLaughlin: assert that there is in fact a proposition that is expressed by the sentence *the singleton set {x} is the sole member of x*, and then deny my claim that I cannot think that proposition, whatever it is. However, how do Fodor and McLaughlin propose to defend their denial of my claim?

There is nothing especially *recherché* about the foregoing example. Other examples come quickly to mind. Chomsky's famous *colourless green ideas sleep furiously* is systematically related to any number of other sentences, some of which share the same syntactic structure, others of which share one or more of this sentence's constituents, but the fact that we can understand (i.e. can think the propositions expressed by) all these other sentences, it does not follow that we can understand this particular sentence. Contrary to what Fodor et al. seemingly suppose, morphosyntactically well-formed sentences constructed out of the constituents of sentences that express propositions need not themselves express propositions.

Now, the foregoing might seem to suggest a possible way of excluding such cases: simply stipulate that the systematicity to be explained holds only for propositions that assert that two things are *symmetrically* related.[6] However, this won't do. As Fodor and Pylyshyn (1988) themselves put it, 'it's not enough just to stipulate systematicity; one is also required to specify a mechanism that is able to enforce the stipulation' (p. 50). Fodor et al. explicitly require that any adequate explanation of systematicity explain what the possession of the capacity for systematically related intentional

[6] Cf. McLaughlin, 1993a, p. 168.

states *consists in*; yet the proposed explanation, which appeals to what McLaughlin calls the 'constitutive basis' of systematically related intention states, offers no explanation or rationale for such a stipulation.

Given their requirement that any adequate explanation of systematicity explain what the capacity consists in, the proposed classical explanation of systematicity is surely specious. It assumes that cognitive capacities are to be explained in terms of certain mental processes defined over mental representations, where the mental representations are assumed to have a combinatorial syntax and semantics, and the mental processes are assumed to be sensitive to the constituent structure of these representations. However, where, precisely, is the explanation? What has to be explained on the classical account is the systematicity of mental representations, i.e. the fact that mental representations exhibit a *particular* combinatorial syntax and semantics—one, for example, in which the capacity to think the thought that aRb and the capacity to think the thought that bRa are systematically related. However, the so-called classical account provides no account of the psychological mechanisms that produce and utilize these representations. The mistake here is akin to that of linguists who having written a grammar for a natural language suppose that in so doing that they have succeeded in explaining the systematicity of natural language. The classical account mistakenly takes the description of the phenomenon for its explanation.

4. *Smolensky's Proposed Explanation of Systematicity: The Classicist's Complaint*

In a series of papers, Smolensky (1987, 1991, 1995) has argued that connectionists can explain systematicity, and furthermore can do so without implementing a classical architecture; in particular, he has argued that they can explain systematicity in terms of vector product representations. The details of Smolensky's proposal are not important here. His argument, very simply, is that (1) connectionist networks are quite naturally viewed as computing functions defined over vector production representations, (2) such representations are fully adequate to express any and all constituency relations expressible by means of classical representations, (3) any and all computational operations definable over classical representations can be captured by connectionist processes defined over vector product representations; hence, (4) connectionists can explain systematicity in terms of such representations.

Fodor and McLaughlin (1990, p. 198) reject Smolensky's argument. Their basic complaint is that the components of tensor product vectors, unlike the constituents of complex classical symbols, don't really exist, and as such cannot be causally efficacious, and hence cannot explain anything:

> 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 . . . is that whereas the Classical constituents of the 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.

The issue, Fodor and McLaughlin (1990) insist, is not whether tensor product representations can *represent* constituent structure; they are willing to assume that they can. Rather the issue is whether tensor product representations have constituent structure, more correctly, whether such representations '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' (p. 200). The answer to this question, they claim, is clear: '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' (p. 200).

Now, coming from Fodor this is a surprising complaint indeed, as Fodor has long defended very liberal criteria for both the existence and causal efficacy of the theoretical entities postulated by special sciences. Roughly, according to Fodor, a thing (kind, property, etc.) exists if a law essential for explaining some phenomenon or capturing some generalization adverts to the thing, and a thing is causally efficacious if a dynamical law adverts to that thing.[7] But by these criteria, the so-called 'normal modes' into which tensor product vectors are typically decomposed and to which Smolensky appeals in his proposed connectionist explanation of systematicity would seem to qualify both as existent and as causally efficacious. For Smolensky (1991, pp. 221–2) argues that vector decomposition into normal modes is essential to explaining regularities in the connectionist network's behaviour:

> To explain the behavior of the system, we usually choose to decompose the state vector into components in the directions of the normal modes, which [determined as they are by the linear interaction equations of the system] are conveniently related to the particular dynamics of this system. . . . There's no unique way to decompose a vector. That is to say, there are lots of ways that this input vector could be viewed as composed of constitents, but normal mode decomposition happens to enable a good explanation for behavior over time. . . .
>
> So, far from being an unnatural way to break up the part of the

[7] Cf. 'P is a causally responsible property if it is a property in virtue of the instantiation of which the occurrence of one event is nomologically sufficient for the occurrence of another', Fodor, 1990, p. 143.

> connectionist state vector that represents an input, decomposing the vector into components is exactly what we'd expect to need to do to explain the processing of that input. If the connections that mediate processing of the vectors representing composite structure have the effect of sensible processing of the vector in terms of task demands, it is very likely that in order to *understand and explain* the regularities in the network's behavior we will need to break the vector for the structure into the vector for the constituents, and relate the processing of the whole to the processing of the parts.

Smolensky, for his part, is all too ready to concede that the normal modes of vectors are causally inert, but the concession is one that Fodor et al. would have done well not to endorse quite so enthusiastically. For what motivates Smolensky's concession is a reductionist intuition that Fodor has taken pains to deny elsewhere,[8] namely, the intution that all the real causal work is being done at a more primitive level of description (1991, p. 222):

> The real mechanism driving the behavior of the system operates oblivous to our descriptive predilection to vector decomposition. It is the numerical values comprising the vector (in the connectionist case, the individual activation values) that really drive the machine.

However, surprisingly, something akin to this same intuition seems also to underpin Fodor et al.'s conviction that classicists, but not connectionists, can explain systematicity. The complaint against tensor product representations, we'll recall, is that they don't actually have constituent structure. They don't have it, because, as McLaughlin (1993a, p. 179) puts it, the normal modes into which the tensor product vectors are decomposed don't 'correspond' to causal agents in the network:

> The subvectors of complex representations won't correspond to causal agents in the network. The causal agents in the network are actual units at various levels of activations. And the constituent representations will not correspond to patterns of activation over units that are actually present in the network.

McLaughlin offers no explication of this notion of 'corresponding' to a causal agent in the network, but the idea seems to be that in order for normal modes to be causally efficacious constituents they would have to stand in something like an isomorphic relation to the actual units of the network, because it is these that are doing the real causal work. It is unclear, incidentally, why on McLaughlin's view anything less than identity would do; isomorphisms are notoriously easily come by and don't necessarily preserve

[8] See, e.g. Fodor, 1990, pp. 138f.

causal relations. However, this difficulty aside, as a general condition on causal efficacy, the correspondence requirement seems certain to consign the entirety of the special sciences, including cognitive science, to a non-causal status. For as Fodor (1974) has persuasively argued, it is a defining characteristic of the special sciences that their taxonomies cross-classify, both with respect to physics and with respect to one another. Such cross-classifying taxonomies are surely incompatible with any correspondence requirement.[9]

Faced with the objection that their position is incompatible with Fodor's liberal criteria for existence and causal efficacy, Fodor et al. might well be prepared to eschew these criteria in favour of more conservative criteria that demand that causally efficacious constituents be physically discrete *parts* of the complex wholes that these constituents constitute. Classical constituents such as figure in the symbol manipulation processes associated with classical computational architectures might typically satisfy these more stringent criteria (but see below); however, there is no evidence, and indeed little reason to suppose, that the constituents of the mental sentences postulated by the language of thought doctrine would satisfy these criteria. Everything that we know about (neuro)physiological realization of computational processes is consistent with the constituents of the complex symbol structures postulated by these processes *not* being physically discrete parts of these structures. To assume otherwise is pure speculation.

It is perhaps worth noting that Fodor et al.'s stipulation that classical constituents are physically discrete parts of the complex symbol structures of which they are constituents imposes a requirement that many of the architectures that Fodor and Pylyshyn (1988, p. 4) themselves identity as 'classical' fail to satisfy. The compression algorithms routinely used in personal computers to compress data files, for example, do not preserve the sort of constituency relation that they envision.[10] Nor does it seem at all plausible to assume that the symbol structures that might figure in connectionist implementations of classical architectures would satisfy the requirement. This may explain McLaughlin's very surprising parenthetical remark (1993a, p. 178) that most classical AI, like connectionist AI, represents syntactic structure only by stipulation—a remark that only underscores how remote the current debate is from issues having to do with computational architectures, as that notion is usually understood.

---

[9] Weinberg (unpublished) presents a more developed version of the foregoing objection. Another way of putting the objection presented here and by Weinberg, one suggested by Egan's, 1995, criticism of attempts to draw eliminativist conclusions from connectionism, is to note that McLaughlin's correspondence requirement imposes a very strong, unsubstantiated constraint on intertheoretic compatibility. As Egan points out, p. 184, very often the complex of structures that realize a causally efficacious state posited at a higher level of theory (her example is the gene that realizes sickle cell anaemia) will be arbitrary from the perspective of a lower level theory. To impose the correspondence requirement is to deny the causal efficacy of such states.

[10] I owe this point to Mark Moyer.

5. Smolensky's Proposed Explanation: A Closer Look

Although Fodor et al.'s objection fails, Smolensky's argument in support of his claim that connectionists can explain systematicity is clearly a non-sequitor. Two computational Doppelgangers (e.g. identical twins perhaps) may share the same systematically related cognitive capacities without the one therefore explaining the other. The problem with Smolensky's argument is that it attempts to draw an epistemological conclusion regarding the availability of an explanation of systematicity from non-epistemological premises regarding the computational capacities of the devices whose systematically related cognitive capacities are to be explained. At very least the argument needs a premise that establishes the explanatory sufficiency of the adduced premises.

The obvious questions here are (1) what will suffice for an adequate explanation of systematicity, and (2) can connectionist models meet this adequacy condition? One well-known answer to the first question, defended most notably by Cummins (1975, 1983), and endorsed by McLaughlin (1993a, 1993b), holds that explanations of cognitive capacities, including presumably systematically related cognitive capacities, must take the form of what Cummins (1975) calls a 'functional analysis'. Roughly speaking, a functional analysis explains a complex molar capacity in terms of the cooperative interaction of certain more primitive capacities that are constitutive of the complex capacity. Thus, for example, David Marr (1982) proposes to explain the molar capacity of the visual system to determine 'what's where' in a subject's immediate visible environment in terms of certain capacities that are said to be constitutive of that capacity, e.g. the capacity of early processing to construct a so-called 'primal sketch' from the deliveries of the retinas, the capacity of later modular processing to compute a solution to the so-called structure-from-motion problem, and so on. McLaughlin (1993a, p. 167) puts the claim that an adequate explanation of systematicity must take the form of a functional analysis this way:

> Classicists take it as a condition of adequacy on a theory of cognition that it explain what possession of capacities to have intentional states consists in. A theory of cognition should describe 'constitutive bases' for such capacities: conditions satisfaction of which constitute possession of the capacities. . . .Classicists thus maintain that an adequate theory of cognition must offer what Cummins (1983) calls a 'functional analysis' of capacities to have intentional states. Systematicity comes into the picture in this way: accounting for systematic relationships among capacities to have intentional states places a substantive constraint on any functional analysis [of] such capacities, that is, on any account of what possession of such a capacity consists in.

The underlying assumption here, as McLaughlin (1993b, p. 219) explains, is that cognitive capacities are *not* fundamental capacities: possession of a cog-

nitive capacity consists in the possession of certain other capacities, the so-called 'constitutive bases' of the cognitive capacity. To explain a cognitive capacity is to specify and describe its constitutive bases, its constitutive capacities. Systematically related cognitive capacities are capacities that are so related, classicists assume, by virtue of *shared* constitutive bases: 'Classicists maintain that the members of the relevant pair of [systematically related] capacities can be functionally analyzed into *common* capacities and (second-order) capacities for their joint exercise' (p. 221). These common capacities *constitute* possession of the members of pairs of systematically related cognitive capacities. And in case the moral for connectionism weren't clear: 'if connectionism does not purport to offer such [functional] analyses, then the issue of explaining systematicity does not arise for connectionism' (p. 225).

It is not at all clear that the demanded form of explanation is going to be available to Smolensky (or to pdp connectionists more generally). In his proposed 'Integrated Connectionist/Symbolic (ICS)' explanation of systematicity, Smolensky (1995) envisions a situation where the input/output functions computed by cognitive processes, at least in core parts of higher cognitive domains, are described by recursive (symbolic) functions. Nevertheless, these functions are *not* computed by means of symbolic algorithms. There may be symbolic algorithms that would compute the functions in question; however, such algorithms do not really describe what is going on computationally: 'In all cognitive domains, cognitive processes are described by algorithms for spreading activation between connectionist units' (Smolensky, 1995, p. 224).

The situation that Smolensky envisions is analogous to that of a parser for English that took sentences as inputs and delivered structural descriptions (trees, labelled bracketings, etc.) as outputs, but that accomplished this feat by computing a function on the natural numbers, specifically one that took the Gödel number of the sentence to be parsed as input and delivered the Gödel numbers of the structural descriptions as output. There would be a symbolic specification of the function computed, perhaps provided by a grammar of the sort proposed by current linguistic theory; there might also be an available symbolic algorithm that could compute this function, perhaps the parsing algorithm implemented by a Marcus parser.[11] But the parser does not in fact execute such a symbolic algorithm. About the only part of the story that such an algorithm gets right is its specification of the inputs and outputs (which are, of course, provably equivalent to the Gödel numbers assigned to those inputs and outputs).

In the situation that Smolensky envisions, the symbolic algorithm provides nothing that would merit describing it as providing a functional analysis of the capacity; after all, the algorithm doesn't explain what, to use McLaughlin's words, the capacity in question 'consists in'. Rather it explains what *in*

[11] Cf. Matthews, 1991.

other devices that very capacity (i.e. for parsing English) might consist in, but not what in this particular device it consists in. However, it is only constituency in the actual case that is pertinent to the explanation. The capacity of the Gödel parser, for example, is the capacity *inter alia* to compute certain functions on the natural numbers, not the capacity to exercise a number of constitutive subcapacities such as building constituent phrases, adjoining them to other constituents, and so on. An explanation of the systematicity of the Gödel parser would mention the former capacity, but never the latter.

The relevant question here is whether even if the symbolic algorithm cannot provide the basis for the requisite functional analysis, might not the pdp algorithm be able to do so. In other words, why can't the connectionist also provide a functional–analytical explanation of systematicity, albeit one that adverts to the individual units of pdp networks?

The individual units are clearly constituents of their respective networks in some sense of that term, and they do have specifiable functions. However, it is not obvious that they would be able to serve the requisite explanatory role. At very least there would seem to be a *problem of grain*: it would be very difficult, if not impossible, to grasp how the molar capacity of the network comprises (consists in) the capacities of the individual units that constitute the network. The specific contributions of individual units would typically be so diffuse as to preclude any claims about which units are responsible for which aspects of the network's molar capacity; moreover, the number of units would typically be so large, their interaction so complex, it would simply be beyond our cognitive ability to grasp how the molar capacity of the network could 'consist in' the capacities of the constituent units. This is not, of course, to deny that the molar capacity of the network does consist in the capacities of the constituent units; rather it is simply to note that one can know the capacities and interactions of all the constituent units and still not be able to understand just how the molar capacity of the network arises out of these capacities and interactions. A proposed explanation of the network's capacity in terms of the capacities of the constituent units (and their interactions) would simply fail to provide the requisite understanding that any such explanation must provide. The difficulty here is epistemological—analogous to that of undertaking to explain the macroeconomic behaviour of, for example, a society in terms of the microeconomic behaviour of its individual members: any understanding of the macrobehaviour would be lost among all the detail about the microbehaviour of the individuals, which is not to deny that the macrobehaviour is an aggregation of the microbehaviours.

Some classicists may think that the foregoing understates the difficulties that face any explanation in terms of the capacities (and interactions) of the network's individual units. The problem, they may think, is not simply that the aggregate microbehaviour of the individual units defies comprehension, but that the individual units are not of the appropriate sort to subserve a cognitive explanation. It may be argued that if an explanation of a cognitive capacity is a truly cognitive explanation, then it must advert only to constitu-

ent capacities that are themselves cognitive in character. Now there are various ways to spell out what is meant here by a 'cognitive' capacity (e.g. that the capacity can be characterized in semantic terms, that the capacity satisfies certain epistemological constraints), but however this is done, the capacities of the individual units in the pdp network will probably turn out to be no more cognitive in character than are the capacities of neurons, synapses, and the like. So whatever else can be said about an explanation that adverts to the network's individual units, it will not be a *cognitive* explanation, and that, of course, is what classicists challenge connectionists to provide.

Part of what is at stake in this further objection is a proprietary dispute over who gets to stipulate what will count as a 'cognitive' explanation. For those who urge this objection, cognitive explanations are distinguished not by their explananda but rather by their explanans,[12] so that explanations that share the same explananda with cognitive explanations need not themselves be such. Cummins seems to endorse such a position in his famous paper (Cummins, 1975), but he disavows or at least modulates the position in the subsequent book (Cummins, 1983). Though in the latter he argues that what he dubs 'interpretive analysis' (either functional analysis followed by componential analysis or componential analysis alone) is *usually* the appropriate explanatory strategy in cognitive science, he points out (pp. 42ff) that explanations of cognitive capacities need not, and sometimes do not, take this form. Some cognitive capacities are 'explicable by instantiation', i.e. by an explanation that adverts to the structure that realizes the capacity (p. 42):

> Not all information-processing capacities require explanation via interpretive analysis. An AND-gate, for instance, is a device with an information-processing capacity that is not subject to analysis into other similar capacities. ... To explain the truth-functional capacity of a (normal) AND-gate does not require, hence does not justify, any further interpretation; what is required is physical instantiation.

Cummins does, however, believe that 'interesting' cognitive capacities will turn out *not* to be explicable by instantiation, but not because he has some proprietary notion of cognitive explanation. Rather he believes that in the absence of an explanatory rationale of the sort provided by a functional analysis, the cognitive capacity will remain 'incorrigibly mysterious to us' (p. 200, n. 8), for reasons that, he concedes, rest on an 'unblushing rationalism' (p. 57).

Cummins' argument is worth scouting here both because it is about the only argument on offer for the claim that explanations of interesting cogni-

[12] Cf. Fodor and Pylyshyn's insistence, 1988, p. 10, that cognitive architecture consists of 'the set of basic operations, resources, functions, principles, etc. . . ., whose domain and range are the *representational states* of the organism' (their emphasis).

tive capacities will (must?) take the form of a functional analysis and because it will provide some measure of the feasibility of a non-classical explanation of the sort Smolensky envisions.

Cummins claims (1983, p. 58) that to understand an 'interesting' cognitive capacity (such as, presumably, the capacity for systematically related cognitive capacities) is to see the system that possesses this capacity as so structured as to exploit whatever it is that makes its output right given the input. However, if, as will supposedly be the case for interesting capacities, our only account of 'what makes a given output right' is that it is derivable from the input via the characterizing rationale that the functional analysis provides, then, Cummins claims, 'such a capacity is explained only on the hypothesis that the system [actually] executes the rationale' (1983, p. 58). In other words, in the absence of an alternative characterization of the cognitive capacity, the explanation of that capacity will have to take the form of a functional analysis. The argument in support of this claim runs as follows (1983, pp. 58–9):

> If the device does not execute the characterizing rationale or any input–output equivalent rationale, then the capacity must be explicable via instantiation (if it is explicable at all). So let's imagine that we have a device with a capacity C that we can only characterize via some rationale [R], and let us consider the hypothesis that C is instantiated as I, a symbol cruncher perhaps, or a physically specified system. This hypothesis commits us to the claim that C and I are isomorphic. What would justify such a commitment?

Cummins argues that nothing short of being able to 'translate' I into R (his terminology), i.e. nothing short of our discovering that 'I is R in disguise' (p. 59), would convince us that C and I were isomorphic. However, if I and R were intertranslatable, then I is, as classicists would put it, a mere implementation of R, so that in instantiating I, the device with capacity C is not simply characterized by R but in fact *executes* R. Thus, the supposed explanation by instantiation turns out to be an explanation by functional analysis.

Cummins' argument, which he dubs his 'what-else' argument, rests on the intuition that in the absence of an alternative to the characterizing rationale (that by assumption one has), there would be no reason to accept as an instantiation of the capacity any process that could not be seen as executing that rationale. This intuition gets expressed in two crucial premises: (i) C's being instantiated as I requires that C and I be isomorphic, and (ii) in the absence of an alternative to the characterizing rationale, nothing short of I's being intertranslatable with R would justify our assuming that C and I are isomorphic. Both the underlying intuition and the premises are false. C's being instantiated as I does not, as Cummins claims, require that C and I be 'isomorphic', whatever exactly that means in the present context; it

requires only that I compute the input/output (I/O) function specified by C. Consider Cummins' own example of the AND-gate, so-called because of its capacity for truth-functional conjunction. Such a capacity can be instantiated in any number of different ways, many of which are not isomorphic to one another (some AND-gates instantiate the function by means of a single primitive operation, while others instantiate the function by means of a logic circuit consisting of OR-gates and connecting inverters). Clearly, if the instantiations are not necessarily isomorphic to one another, then they are not necessarily isomorphic to the capacity that they instantiate, since by all accounts isomorphisms are symmetric, transitive relations. Now once we see that instantiation does not demand isomorphism, the problem with the second premise and hence with the underlying intuition becomes apparent: there is no reason to suppose that in the absence of any alternative to the characterizing rationale (that by assumption one has) one could be justified in concluding that C is instantiated as I *only if* I and R were intertranslatable. It will be enough if the instantiation I computes the I/O function specified by R. However, if that is all that is required, then the mere fact that R provides a characterizing rationale for C, even the sole such rationale, would provide no reason to suppose that a device with capacity C actually executes R. Nothing in all this provides any reason to suppose that explanations of cognitive capacity must take the form of a functional analysis, even in the situation where we lack any alternative to a characterizing rationale for the capacity.

While Cummins' 'what-else' argument fails, it does suggest a different line of argument which, if successful, might establish that explanations by instantiation are not likely to be available for 'interesting' cognitive capacities, so that if such capacities are to be explained, then the explanation will most likely have the form of a functional analysis. The argument runs as follows. It is a consequence of a formal undecidability result known as Rice's Theorem that there is no effective procedure for deciding whether two arbitrarily chosen partial recursive functions are extensionally equivalent. Thus, while it may be enough, as we argued above, that I computes the I/O function specified by some characterizing rationale R, there is no effective procedure for deciding for some arbitrary R whether some arbitrary I computes the I/O function specified by R. So even though something less than intertranslatability will do, there is no effective procedure for determining whether this weaker criterion is satisfied and hence for determining whether I in fact instantiates the capacity C, which is the object of explanation. The absence of an effective procedure does not, of course, preclude a determination in every case: in many cases it will be apparent simply by inspection whether or not I computes the I/O function specified by R. But in many other cases, most especially in cases of 'interesting' cognitive capacities whose associated characterizing rationale R and instantiation I are likely to be very complex, it will not be so apparent. In those cases, the weaker criterion of extensional equivalence may not be usable, and it isn't obvious that there is any other criterion available, except the criterion of intertranslat-

ability proposed by Cummins. So we might very well be left with Cummins' conclusion, namely, that there can be no explanation by instantiation of interesting cognitive capacities.

This argument based on Rice's Theorem assumes, as did Cummins' 'what-else' argument, that the I/O function computed by the cognitive capacity to be explained is antecedently given (by the specification of R). It further assumes that the task facing someone who proposes an instantiation explanation of the capacity is (therefore) that of ascertaining whether the instantiation hypothesized by the explanation in fact computes the given function. Both assumptions are false. Cognitive capacities don't come labelled by a characteristic I/O function. What function characterizes a cognitive capacity is a matter for empirical discovery. Both the rationale R and the instantiation I are hypotheses, possibly competing, about the proper characterization of the function computed by the capacity in question. As such, the empirical task is not therefore one of discovering an instantiation that computes an antecedently given characterizing function; rather the task is simply to provide an instantiation characterization of the capacity, one that explains the capacity to produce a certain output (given a certain input).

So here is where things stand. Smolensky has given us no reason to think that, as he claims, connectionists can explain systematicity. His argument in support of that claims fails. If explanations of systematicity must, as classicists claim, take the form of a functional analysis, then connectionists are indeed in trouble for the epistemological reasons mentioned. However, nothing we have seen establishes that such explanations must take this form. Certainly neither Cummins' 'what-else' argument nor the argument based on Rice's Theorem provide any support for such a conclusion. So the question remains whether connectionists can explain systematicity, though in a manner other than by way of a functional analysis. I am cautiously optimistic that they can. By way of a conclusion, let me try to justify this optimism.

6. A Connectionist Recipe for Explaining Systematicity?

The challenge to connectionists is to explain the capacity for systematicity without appealing to a characterizing rationale of the sort that a functional analysis would provide and a connectionist device might execute (since to explain the capacity in this way would be to implement a classical architecture). The explanatory task facing connectionists is twofold. First, there is the task of specifying, in a way that is genuinely explanatory, a computational architecture that, under the appropriate intentional interpretation, constitutes the capacity for systematically related intentional states; and second, there is the task of providing the requisite intentional interpretation.

If we put to one side, as classicists have so far been willing to do, the failure of connectionists to provide a computational construal of intentional

states for connectionist architectures, the chief obstacle facing a connectionist explanation of systematicity is that of establishing the explanatory *bona fides* of connectionist architectures. The formal results mentioned earlier in the paper establish that connectionist devices can, at least in principle, exhibit systematicity; however, it remains an open question whether, if exhibited by a connectionist device, this capacity could be explained. Connectionist architectures are generally not amenable to functional analysis, and this, classicists argue, is the only available form of explanation.

Connectionists, I believe, can offer a type of explanation by instantiation, i.e. an explanation that adverts to the structure that instantiates or realizes the capacity to be explained. However, the explanation differs importantly from the instantiation explanations that are typically adduced by Cummins as examples of such explanations. In the examples he adduces, one explains the capacity of, for example, an AND-gate, by describing the structure of the AND-gate that realizes, and thus is causally responsible for, the device's capacity for truth-functional conjunction. The instantiation explanations that connectionists can offer of systematicity do not describe the realizing structure of the particular device whose capacity for systematically related intentional states is to be explained. Rather one explains the capacity in question more indirectly, by something akin to an inductive argument from simple cases. More precisely, the instantiation explanation proceeds as follows. First, the capacity to be explained is (re)described in terms that enable the explanation to establish that this capacity is possessed by certain very simple devices for which the possession of the capacity is explained in the direct way that Cummins envisions instantiation explanations to proceed. Next, the explanation establishes that the complex device whose capacity is to be explained shares with these simple devices the structural properties that in the simple devices explains their possession of the capacity. Then, barring any specific reasons for thinking that the instantiation explanation of the simple cases cannot be extrapolated to the complex case, the instantiation explanation for the simple cases together with the extrapolation theory (i.e. the theory that provides both the (re)description of the capacity and the specification of the shared structural properties) is taken to constitute an instantiation explanation of the complex device's possession of the capacity, even in the absence of a direct instantiation explanation of that device's possession of the capacity.

The sort of instantiation explanation that I am describing here is actually very common; it is arguably the canonical form of engineering and biomedical explanation, where a more direct explanation is often precluded by the complexity of the system whose behaviour or capacity is to be explained. Structural engineers, for example, explain the dynamic instability of certain suspension bridges under transverse wind loadings (e.g. of the ill-fated Takoma Narrows 'Galloping Griddy' Bridge) only indirectly: they provide a theory of the behaviour, under dynamic loading, of unsupported flat sections that are pinned at either end, then extrapolate this explanation to include the cases to be explained. In rough terms, the explanation runs as follows:

unsupported flat sections that are pinned at both ends and that possess structural properties P_i are dynamically unstable under dynamic transverse loadings L_j; the suspension bridges whose instability is to be explained share these properties, and hence are dynamically unstable. Cancer researchers similarly explain the capacity of the male sex hormone testosterone to promote (or in some cases suppress) metastatic prostate cancer: they describe the specific mechanisms by which the hormone promotes (or suppresses) the growth of human cancer cells grown in laboratory dishes and implanted in mice, then extrapolate this explanation to the human models that are the focus of their explanatory interests.

For the case at hand, i.e. for the connectionist explanation of systematicity, the explanation would begin by first providing a formal characterization of systematicity, the capacity to be explained. Specifically, one would characterize it as the capacity to compute any of a set of partial recursive 'systematic' functions, i.e. functions each of which has as its domain, formal objects (sentences, representations, etc.) that are systematically related. In the case of our capacity for natural language, which is by all accounts paradigmatically systematic, the functions in question would simply be the grammars (or parsers) for those languages (however specified), or perhaps a learning function that takes as input the experience on the basis of which a child acquires his native language and gives as output a grammar (or parser) for that language. The explanation would then appeal to the formal result, mentioned earlier in the paper, that connectionist devices can approximate, to any arbitrary degree, any function computable by a classical device. An understanding of the proof of this formal result, along with the suggested formal characterization of systematicity, would (subject to an important qualification discussed below) constitute a connectionist explanation of systematicity of the indirect instantiation type described above. The proof does not simply establish that connectionist devices have a general capacity, one that the accompanying formal characterization establishes to properly include the capacity to be explained; it also explains, by what I'm calling indirect instantiation, why (or how it is) that they have this capacity.

There are a number of important details that remain to be filled in, not to mention the matter of the computational construal of intentional states that we put aside. However, even if we neglect these matters, the proposed instantiation explanation remains importantly incomplete (and this is the qualification mentioned above). Nothing yet explains why, seemingly as a matter of nomological necessity, cognitive systems have the capacity in question. If the proposed explanation is to succeed, then there will have to be some principled way, amenable to explanation, of constraining the class of partial recursive functions available to the connectionist devices that are said to model those systems that exhibit systematicity. Classicists, we have seen, propose to achieve this constraint by demanding that systematically related objects share in common certain 'constitutive bases'. Classical (symbolic) AI models of cognition sometimes enforce these constraints in a more direct fashion, simply by encoding in the model's algorithm an explicit statement

of the constraints in question.[13] Connectionists have no similar means available to them; they must find other, less direct means. Constraining the training set used to induce a certain function clearly won't do, because the relevant constraints are presumably innate. Constraints on the class of available functions would presumably have to be enforced by fixing or limiting the range of the connection weights and/or the activation values of the individual nodes in the network.

There are, so far as I know, no concrete proposals on offer for how to go about constraining multilayer networks so as to limit the class of available functions to only systematic functions (or some subclass thereof), but the prospects may not be as bleak as classicists have supposed. Connectionists working within a so-called 'structured connectionist' approach have recently had modest success in modelling some of the innate constraints on natural language, constraints that effectively guarantee that the class of functions computable by connectionist devices incorporating these constraints will be systematic.[14] Other connectionists have begun to pursue formal learnability problems within a connectionist paradigm, examining the learnability (under various constraints on access to data, memory, and the like) for particular classes of formal grammars.[15] The results thus far are limited (as, incidentally, were analogous results in the classical paradigm until relatively recently), but they are encouraging. If these limited successes can be extended, and also achieved in other cognitive domains, then connectionists might very well find themselves with a principled way, amenable to explanation, of constraining the set of functions available to simple multilayer networks to only systematic functions. Such a scenario seems quite plausible, because the innate constraints that are enforced in a given cognitive domain seem to track pretty well the conception of systematicity peculiar to that domain. The explanation of how these innate constraints enforce systematicity in simple networks might in turn provide the basis for an indirect instantiation explanation of how systematicity is enforced in the more complex networks that would presumably be offered as models of human cognition.

Department of Philosophy
Rutgers University

References

Chalmers, D. 1990: Syntactic Transformations on Distributed Representations. *Connection Science*, 2, 1, 2, 53–62.

---

[13] It is noteworthy that proposed classical models of parsing and vision very often achieve these constraints in more indirect ways, see, e.g. Matthews, 1991, pp. 185–92.

[14] See, e.g. Regier, 1996, as well as references therein to structured connectionist research.

[15] See, e.g. Kremer, 1996.

Cummins, R. 1975: Functional Analysis. *Journal of Philosophy*, 72, 20, 741–64.
Cummins, R. 1983: *The Nature of Psychological Explanation*. Cambridge, MA.: MIT Press.
Egan, F. 1995: Folk Psychology and Cognitive Architecture. *Philosophy of Science*, 62, 2, 179–96.
Fodor, J. 1974: Special Sciences. *Synthese*, 28, 77–115.
Fodor, J. 1975: *The Language of Thought*. New York: Thomas Crowell.
Fodor, J. 1987: *Psychosemantics*. Cambridge, MA.: MIT Press.
Fodor, J. 1990: *A Theory of Content and Other Essays*. Cambridge, MA.: MIT Press.
Fodor, J. and McLaughlin, B. 1990: Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work. *Cognition*, 35, 2, 183–204.
Fodor, J. and Pylyshyn, Z. 1988: Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28, 3–71.
Hornik, K., Stinchcombe, M. and White, H. 1989: Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, 2, 359–66.
Kremer, S. 1996: A Theory of Grammatical Induction in the Connectionist Paradigm. Unpublished PhD dissertation, Department of Computing Science, University of Alberta, Edmonton, Alberta.
Marr, D. 1982: *Vision*. New York: W.H. Freeman.
Matthews, R. 1991: Psychological Reality of Grammars. In A. Kasher (ed.), *The Chomskyan Turn*. London: Basil Blackwell, 182–99.
Matthews, R. 1994: Three-concept Monte: Explanation, Implementation and Systematicity. *Synthese*, 101, 3, 347–63.
McLaughlin, B. 1993a: The Connectionism/Classicism Battle to Win Souls. *Philosophical Studies*, 71, 163–90.
McLaughlin, B. 1993b: Systematicity, Conceptual Truth, and Evolution. In C. Hookway and D. Peterson (eds), *Philosophy and Cognitive Science*, Royal Institute of Philosophy, Supplement No. 34, 217–34.
Regier, T. 1996: *The Human Semantic Potential*. Cambridge, MA.: MIT Press.
Smolensky, P. 1987: The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn. *Southern Journal of Philosophy*, 26 (suppl.), 37–63.
Smolensky, P. 1991: Connectionism, Constituency, and the Language of Thought. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and His Critics*. London: Basil Blackwell, 201–227.
Smolensky, P. 1995: Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Cognitive Architecture. In C. MacDonald and G. MacDonald (eds), *Connectionism: Debates on Psychological Explanation*, London: Basil Blackwell, 223–90.
Weinberg, J. (unpublished): Causation and Explanation in the Connectionist/Classicist Debate.