Directions in Connectionist Research: Tractable Computations without Syntactically Structured Representations

Jonathan Waskan and William Bechtel
Philosophy-Neuroscience-Psychology Program
Washington University in St. Louis

When connectionism reemerged as a framework for modeling cognitive processes in the 1980s[1], it was a relatively well defined approach. One of the essential components of this reemergence was the discovery of the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986), which overcame what had seemed to be an insuperable problem for one of the major approaches to connectionism during its previous period of prominence, Rosenblatt's perceptron model (1962; see Minsky and Papert, 1969, for the objections that seemed to show the in-principle inadequacy of the connectionist approach). With the availability of the backpropagation learning algorithm, the prototype of a connectionist network became the three-layer feed-forward network (see Figure 1) in which a problem is supplied to the network in the form of a pattern of activation values on the input units and the network generates an answer in the form of a pattern of activation values on the output units. This is accomplished by a set of weighted connections in terms of which each unit in one layer influences the activation value of a unit in the next layer; it is the weights on the connections which are acquired through the backpropagation of error messages through the network whenever it produces incorrect outputs.

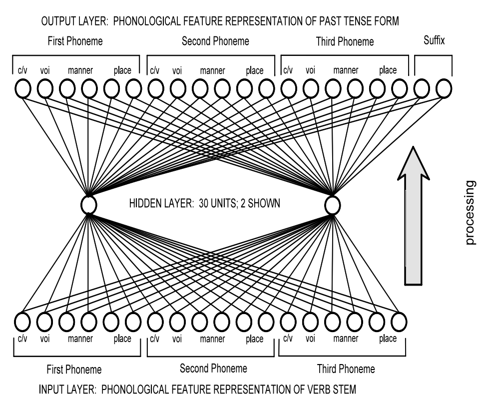

Figure 1: A prototypical example of a three-layer feed forward network, used by Plunkett and Marchman (1991) to simulate learning the past-tense of English verbs. The input units encode representations of the three phonemes of the present tense of the artificial words used in this simulation. The network is trained to produce a representation of the phonemes employed in the past tense form and the suffix (/d/, /ed/, or /t/) used on regular verbs. To run the network, each input unit is assigned an activation value of 0 or 1, depending on whether the feature is present or not. Each input unit is connected to each of the 30 hidden units by a weighted connection and provides an input to each hidden unit equal to the product of the input unit's activation and the weight. Each hidden unit's activation is then determined by summing over the values coming from each input unit to determine a netinput, and then applying a non-linear function (e.g., the logistic function $1/(1+e^{netinput})$. This whole procedure is

[1]Seminal events that marked its reemergence were the publication of Hinton and Anderson (1981), the Cognitive Science Society meetings in 1984, and the publication of the PDP volumes (Rumelhart, McClelland, and the PDP Study Group, 1986, McClelland, Rumelhart, and the PDP Study Group, 1986).

Three-layer feedforward networks did much to change the conception of explanation in cognitive science. For example, they made pattern recognition (as contrasted with reasoning by means of applying formal rules) prototype of a cognitive process. They also showed how cognitive processes might make use of soft-constraints, operate with incomplete representations, and learn from previous experience. They also fostered new ways of thinking about mental representations by introducing notions such as distributed representations and coarse coding. These features of connectionist networks became the focus of many philosophical discussions of connectionism (see Bechtel and Abrahamsen, 1991, Clark, 1989, 1993, and papers in Horgan and Tienson, 1991, and Ramsey, Stich, and Rumelhart, 1991).

As connectionist modeling took root in cognitive science, though, it also diversified. Researchers have explored different architectures (e.g., recurrent, interactive, as well as feedforward), different learning algorithms (e.g., variations on Hebbian unsupervised learning), and different, more complex tasks (controlling robots and reading and paraphrasing stories). Of particular note has been the introduction of tools of dynamical systems theory to analyze the behavior of networks. In our discussion that follows, we will describe examples from the connectionist literature of some of these new developments in connectionist modeling.

Pursuit of diversification has corresponded to a general acceptance of connectionism as a framework in which to model cognitive processes. As this has happened, some of the controversial issues of the 1980s, especially the conflict between classicism and connectionism, have largely faded from view. But sometimes it is useful to return to the old issues, for we may be able to see them in a new light and this may help us to evaluate the significance of current contributions. Terence Horgan, in the previous paper, and Horgan and Tienson (1994, 1996) have brought back the controversy between classicism and connectionism in just such a useful way. They identify two sets of issues on which classicism and connectionism might differ. The first concerns the assumption that one can identify rules that operate on representations and that, in terms of such rules, one can explain cognitive processes in terms of "tractably computable cognitive transition functions." Horgan argues that connectionism ought to split with classicism on this issue. The other concerns the use of syntactically structured representations. Horgan claims that the common interpretation of connectionism (which he terms "non-sentential computionalism") differs from classicism primarily in rejecting this assumption. He contends that this is a mistake. Instead of strengthening the inadequate resources classicism offers for modeling cognition, Horgan contends that this move depletes the resources further, resulting in a "seriously crippled cousin of classicism" (ms p. 17). He graphically portrays what he takes to be the folly of the move:

> What exactly are we supposed to be *gaining*, in terms of our abilities to model cognitive processes, by adopting an approach which (i) retains the assumption that cognitive processing is representation-level computation, but (ii) eschews one extremely powerful way to introduce semantic coherence into representation-level computation: viz., via the syntactic encoding of propositional content? Qua representation-level computation, it looks as though this amounts to trying to model thought processes with one hand--the *good* hand--tied behind one's back (ms. p. 17).

We contend that Horgan has made the wrong call on both of these issues. That is, we claim that cognitive science is and should be engaged in seeking tractable representation-level computational processes (although not at the level of whole cognitive transition functions) and, while these computations are over representations, we argue that these representations should not be required to be syntactically structured.

1. Precise, exceptionless rules and tractably computable transition functions

Those who revived connectionism were led back to it out of a perception that the classical approach to cognitive modeling faced serious short comings. One of these was that classical models were excessively brittle. If an AI system were given a novel, unanticipated input, were overloaded, or if it was partly damaged, it would simply fail to operate, whereas human cognitive agents generally offer intelligible responses to new inputs, do not simply crash when faced with cognitive overload, and continue to perform after moderate neural loss. (For further discussion, see Bechtel and Abrahamsen, 1991.) Horgan identifies a different failure of classical approaches to cognitive modeling, the long-standing frame problem. The term "frame problem" stems from McCarthy and Hayes (1969), who identified it as a problem for the deductivist approach to AI. The problem consists in specifying in rules both what changes and what stays the same when an action is performed. It seems to be impossible to state the axioms that a system would need in order to update each piece of information in its working memory after any possible action. Since its introduction, the frame problem has been generalized as a problem of how a system could maintain an accurate representation of the world in which it acts (for a useful review, see Haselager, 1995).

Horgan follows Fodor (1983) in linking the frame problem with the problem of non-demonstrative inference or empirical confirmation in science; Fodor contends non-demonstrative inference is problematic because it is isotropic and Quinean. It is isotropic since, in principle, anything that scientists know may be relevant to evaluating a new proposition; it is Quinean since assessing whether to believe a new proposition depends on features such as coherence and simplicity of the whole web of a scientist's belief. For Horgan, the root of the problem for classicism is the insistence on solving the problems with *tractable* computation. Thus, with respect to the problem of considering features of the whole web of belief in deciding whether to believe a new proposition, Horgan says:

> Not only do we have no computational formalism that show us how to do this; it's a highly credible hypothesis that a (tractable) computational system with these features is just impossible, for belief systems on the scale possessed by human beings (ms., p. 9).

The promise of finally solving the frame problem is an exciting prospect for connectionism. Horgan's proposal is that his preferred version of connectionism, dynamical cognition, can solve the problem, at least in its Fodorian guise, by rejecting the appeal to precise, exceptionless rules and tractably computable transition functions. In this section, we shall offer three sorts of response: (a) we question whether giving up the quest for tractable computation through rules operating on representations solves the problem; (b) we challenge the contention that human cognition is as isotropic and Quinean as Fodor and Horgan assume; and (c), we argue that current progress in cognitive science, especially cognitive neuroscience and connectionist modeling of cognitive processes is stemming from continuing the quest for identifying component systems performing tractable computations.

Before we begin, there is a feature of Horgan's presentation that we need to take up. In his exposition of the condition of tractable computation, Horgan speaks of total cognitive states, which he later analyzes in dynamical systems' terms as points in the multidimensional space defined over the parameters used to describe the system (in connectionist terms, activation values). While it is clear that the dynamical systems' approach of focusing on transitions in total dynamical systems can sometimes be very useful, the systems that are analyzed this way by dynamicists are much simpler systems than the total cognitive system (for examples, see Kelso, 1995, Beer, 1995, Thalen, 1995). For purposes of analysis, one can always combine coupled dynamical systems into one system, but the guiding question is whether it is useful to do so. Moreover, for purposes of tractable computability, the focus on total cognitive states is misguided for both classicism and connectionism. Both approaches tend to focus on specific cognitive tasks, and design systems for these. One might use a three-layer feedforward network to model specific tasks, but no one has taken seriously the idea that the whole cognitive system should be modeled as such a network. Likewise, one might devise a production system or a case-based reasoning system to handle specific tasks, such as story comprehension, but this is recognized to be a model of a part of the cognitive system. When Newell (1989) advocated SOAR as a unified theory of cognition, he was advocating that the production system architecture used in SOAR would be the basis for all cognitive modeling, not that one SOAR system would handle all cognitive activities. Only in the robotics field is there a concern for devising a classical or connectionist system to handle the whole cognitive life of an agent, but the cognitive lives of these agents is extremely restricted. Nearly everyone in cognitive science assumes that modeling a complete cognitive agent will involve the integration of many systems (e.g., production systems or connectionist networks) handling specific tasks and sharing only partial information. The interactions of these systems may, and likely will, turn out to be nonlinear, making the overall cognitive transition function non-tractable[2]. But this is not taken to be a problem since tractable computation is sought by both classicists and connectionists in the performance of individual tasks.

a. Does giving up tractable computation solve the Quinean and isotropy problems?

The picture that Horgan advances is that the requirement of identifying rules operating on representations and tractable computations excessively limits the mathematical relations that one could employ in explaining cognitive performance. Without these limitations, Horgan seems to be implying, one might hope to account for the isotropic and Quineian nature of cognition. It is not clear to us how allowing for a larger set of mathematical tools will explain the isotropic and Quinean nature of cognition. The problems of isotropy and the Quinean nature of cognition are ones of getting to the *right* information in the system and measuring the *appropriate* characteristics of the whole network of beliefs (e.g., its consistency). These are not just (and may not even be questions of) limitations imposed by keeping the computations tractable, rather, these are problems about how to organize information in a system.

Horgan and Tienson (1994) do suggest three features of connectionist systems which may allow them to address the isotropic and Quinean character of cognition: that knowledge is contained in the weights and not explicitly represented, learning new tasks involves changes in weight settings, which alter the response characteristics of the network, and that processing can involve satisfaction of soft-constraints. We agree that these are all important features of connectionist networks. We question the suggested connection between these features and the rejection of tractable computation, the claim that these are exclusively features of connectionist networks, and whether they are sufficient to give networks isotropic and Quinean characteristics. With respect to the first point, it is important to note that these are features of connectionist networks that have been emphasized repeatedly since the reemergence of connectionism and those emphasizing these features were not

---

[2]Another crucial issue raised in Horgan's presentation is the notion of tractability. What makes a computation non-tractable? In part, Horgan and Tienson illustrate what non-tractable computation would be in terms of a huge (possibly infinite) *list* of functions for which there is no more compact representation. They do not say what makes a list huge, so one cannot determine whether, for example, the rule set used in SOAR is tractable.

focused on denying tractable computation in connectionist networks. They are present, moreover, even in simple networks such as two-layer networks with linear activation rules. It is the case that once you have distributed computation performed with multiple components interacting in a non-linear manner, the input-output relation will become virtually intractable; one may have to calculate separately the output function for each sending unit, the netinput to any receiving unit, and the consequent activation function for each unit in the system in order to determine its behavior.[3] This would seem to be the sort of computation that Horgan construes as non-tractable. But the advance lies in these features of networks themselves, not in connectionists having relinquished any commitment to tractable computation.

The second thing to note about these features is that they are not exclusively features of connectionist networks: they are found in many classical systems as well. As Horgan himself notes, classicists are not committed to all information being explicitly represented. Some of it may be hardwired. But more importantly, classical systems often include something very much like weights. In production systems such as ACT and SOAR there are often numeric parameters which can be adjusted as systems learn from experience and which affect the likelihood of a rule firing in a given context (Anderson, 1983, 1993; Rosenbloom, Laird, Newell, and McCarl, 1991). Such parameters often figure in the conflict resolution process, a process very reminiscent of constraint satisfaction. Moreover, an important part of some classical systems is a semantic memory (Quillian, 1968), which contains labeled links (is a, part of, etc.) between different semantic entries and which employs a process of spreading activation to activate new items and add them to working memory. Interestingly, including these features in production systems may serve to make them non-tractable in just the way Horgan portrays connectionist systems. One has to trace the processing through the simulation--there is no simpler manageable set of rules linking inputs to outputs.

The other question that must be addressed, though, is whether any of these features is sufficient to provide for the isotropic and Quinean character of cognition. Horgan and Tienson (1994) point to the fact that weights in networks acquired during learning have the tendency to reactivate representations previously produced in response to one input when similar inputs are provided. This, they say, may help explain the phenomenon of relevance addressable memory, which is one aspect of the problem of isotropy. The question, though, is whether such weights will access all of the information relevant to a problem. Unless the training set has been extremely fortuitous in preparing the network for the new problem, it would not. Accessibility of information stored in the weights of a network is dependent on the way the network has been trained. Unless it has been trained in just the right way, a network will not be able to summon representations developed in the context of another task even if they are relevant to the one before it (Clark & Karmiloff-Smith, 1994).

---

[3] In one respect, things are not as bleak as this makes it seem. Connectionists have developed a variety of tools for analyzing the performance of networks, such as cluster analysis and principal components analysis (see Elman, 1991). These are statistical techniques which enable one to identify the major determinants of behavior (e.g., the principal component that is most responsible for a network keeping track of information relevant to one of its tasks). However, there are limitations to this approach: by collapsing dimensions, for example, one explicitly throws away information. This information may be noise, but it may also be critical to some of the responses the network gives or would give to a new input. Thus, while these are useful tools for understanding network behavior, they do not solve the intractability problem Horgan is addressing.

> Second, Horgan and Tienson focus on the fact that weight changes alter the whole system: When a representation is re-activated after weight changes, it may be recreated with suitable modifications determined by the new weights. This suggests an answer to the problem of updating memory when new information is acquired: there is nothing in memory to be updated; nothing is changed except the weights, and weight changes bring it about that representations are appropriately modified when and if they are re-activated (ms., pp. 28-29).

Of critical importance here is the claim that modifications to a network will be appropriate. Changing weights to solve one problem while sacrificing the ability of a network to solve other problems on which it had been trained previously has been a serious challenge for connectionists. Far from exhibiting proper updating in the wake of new information, many networks exhibit catastrophic interference (McCloskey and Cohen, 1989; Ratcliff, 1990) when they are trained on new tasks without being retrained on old tasks. (One of the more plausible solutions to this problem is to incorporate modularity into networks, in a manner we shall discuss below.) Related remarks apply to soft-constraint satisfaction in networks. This involves allowing conflicts between two possible responses to a problem to be resolved by a competition procedure (realized, for example, by providing inhibitory connections between the two possible responses). The procedures that networks (and production systems) employ to resolve conflicts do have the result that they will always give a response to a new problem that is partly like two or more old problems, but these procedures do not guarantee that the network will behave rationally in its response. Whether it will do so or not will be determined by whether the training set has given the network the right set of weights to do so.

In general, then, if networks succeed in exhibiting the Quinean and isotropic characteristics Horgan addresses, most of the success will be due to the nature of the training set to which the network was exposed (or to the initial weights assigned to connections prior to training). Moreover, the features that are most pertinent to explaining their performance, for example including information in the weights, are features shared by some classical systems. Finally, surrendering tractable computability is not what will enable networks to deal with the Quinean and isotropic character of cognition, rather, non-tractability is a consequence of other design features of the system and may also arise in some classical systems.

b. Is cognition isotropic and Quinean?

Recall that Fodor appealed to the problem of confirmation in arguing for the isotropic and Quinean character of cognition. It is important to note that the problem of confirmation in philosophy of science has traditionally been a normative problem: *should* a scientist believe a new experimental finding or a new theory? Assuming that scientists seek consistent sets of true beliefs that are explanatorily and predictively powerful, the question becomes one of identifying a procedure that would evaluate the degree of confirmation of each proposition. Such a logic of confirmation would then constitute a normative standard against which human performance could be measured. The fact that such a logic of confirmation has not been produced does not have any consequences for modeling cognition unless we have reasons to believe that humans act in such a normatively appropriate way.

We have, up to now, granted Horgan's contention that cognition is isotropic and Quinean. We are not convinced, however, that either ordinary belief fixation or belief fixation through scientific reasoning are as isotropic or Quinean as Fodor and Horgan suggest. Horgan is correct in asserting that a system which is isotropic and Quinean in the strictest sense will prove nearly impossible to model on any large scale. Such a system will need to incorporate what he terms 'relevance addressable' memory and a means of *rapidly* assessing the degree of simplicity and conservatism achieved through the "vastly many competing, incompatible, ways of revising the whole system" (ms., pp. 9). While humans often draw upon a wide range of information in the fixation of a particular belief, it is not obvious that we ever consider *all* of our existing relevant information. Likewise, the degree of confirmation of a particular belief may often be a function of non-local factors such as simplicity and conservatism amongst a larger set of beliefs (ms., pp. 8), but it is not obvious that simplicity and conservatism are ever attained on a truly global scale.

Consider first that any system which rapidly achieves optimal levels of global simplicity through conservative revisions of belief would have a dramatic advantage over your typical human cognizer. Such a system would have little need for either philosophers or theoreticians. Philosophers and theoreticians are, by trade, seldom charged with the responsibility of gathering new information. The primary role of these individuals (as evidenced by the present debate) seems to be the *painstaking* search for ways in which beliefs may be organized in an optimal manner. Often the proposals of such individuals are rejected not because *new* relevant information comes to light, but because beliefs already shared by both proponent and antagonist are shown to be relevant. The mere fact that revisions amongst sets of beliefs are not limited to occasions when new information is being evaluated for possible incorporation into the larger system suggests that human cognition lacks the strict isotropic and Quinean character Fodor and Horgan would have us attribute to ourselves. A *strictly* isotropic and Quinean system would not need to engage in this ongoing reflective activity.

The failure of human cognitive processes to meet the strict isotropic and Quinean demands is also (and perhaps more markedly) evidenced amongst lay-persons. Consider, for example, a common experience of students who are first introduced to traditional philosophical problems such as the freedom/determinism debate. Students often discover that they believe both that humans (themselves in particular) are free and that the universe is deterministic (at least in a sense that precludes the possibility of their own freedom). Becoming aware of this tension often motivates a reorganization amongst previously unchallenged beliefs. For instance, one might follow A. J. Ayer and embrace a weakened version of freedom that is compatible with strict determinism; one might reject either freedom or determinism outright; or, perhaps, upon exposure to Kantian thinking, one might adopt a perspective which seems to enable both freedom and determinism to be maintained in undiluted form. Whatever resolution one prefers, the need for a shift in belief structure is normatively mandated when, and only when, one becomes aware of two previously unchallenged yet incompatible beliefs. Importantly, the process of searching for a new organization is not motivated by a new belief; rather, it results from recognition of a pre-existing incoherence amongst beliefs which should not exist if, on integrating every new belief, cognizers had access to all other beliefs which are relevant to the confirmation of the belief in question. Likewise, the process of achieving this new structure is neither unconscious (see Fodor, 1983, p. 104) nor rapid (see Horgan, ms., p. 9), and it is often accomplished with outside assistance, e.g. from experts in the relevant discipline.

Insofar as the most appropriate model of human reasoning is one in which simplicity and conservatism are salient properties, such properties may be best understood as a kind of implicit ideal which is never attained but is sought in an ongoing, *non-rapid*, and often conscious manner. More often than not the bulk of the task of organizing beliefs is not carried out by the individual at all; rather, most of the organization pre-exists in the form of scientific paradigms and belief systems. Despite the ideals of some pedagogical philosophies, individuals generally adopt the belief systems of their culture and do not, figuratively speaking, reinvent the wheel. One might argue that paradigms are themselves the product of creative genius on the part of scientists. Admittedly, a foundational insight for a new paradigm may be the product of one individual. However, it is seldom the case that a single individual has the time or intelligence to reorganize an entire body of findings into a coherent system; rather, such work is typically accomplished piecemeal by a large number of individuals.

Insofar as the integration of new information is concerned, it may be that (particularly in the scientific realm) content addressability is sufficient to enable a large degree of integration with other beliefs to take place. But even in science, advances occur when someone finally recognizes the relevance of already known information. Once it is admitted that cognitive systems do not access *all* relevant information bearing on the confirmation of a new belief, the possibility arises that the content of a particular belief is sufficient to (over the course of lengthy consideration) enable one (particularly a specialist) to bring to awareness relevant beliefs to the extent that would be expected given the imperfect degree of global consistency that human cognizers seem typically to possess. If this account of human reasoning processes (according to which much of the organization amongst beliefs is pre-existing and the level of global consistency is always imperfect) is correct, then the task of modeling them should seem far less daunting (as compared with the difficulty of modeling on the large-scale a strictly isotropic and Quinean system) for both classicists and connectionists alike.

c. Progress via tractable computation over representations in component sub-systems.

Surprisingly little attention within cognitive science has been devoted to the problem of determining the properties of belief systems. Where it has dealt with issues of encoding and retrieval in semantic memory it has typically worked at a lower level, such as memory for individual concepts. While little attention has been devoted to the process of belief fixation, many other exciting avenues of research at both the physical and abstract functional level have been pursued with respect to, for example, processes underlying attention, perception, pattern recognition, implicit and explicit long term memory, visual and phonological short term memory, and lexical and conceptual representation. One guiding assumption of cognitive neuroscience is that the identification of distinct systems at the functional level of description may be paralleled by identification of distinct systems at the physical level, i.e. that functional process dissociation will be reflected by distinct systems at the level of implementation. Numerous ablation, imaging, and neuro-anatomical studies have been aimed at localizing such systems to discrete neurological structures.[4] Progress has typically been made in cognitive science through identification of the systems subserving a given processes at the levels of both function and implementation. As illustrated below, analysis of such systems (i.e. filling in details at the algorithmic level) is typically accomplished through a description of the kinds of information such systems represent and the manner in which systems process such representations, that is, in terms of tractable computation over representations.

---

[4]Occasionally the need to posit a distinct system underlying a particular process is challenged. For example, Roediger (1990) challenges the utility of the distinction between implicit and explicit memory. Relatedly, Hintzman (1986) challenges the need for separate systems subserving the encoding of episodic and conceptual information. In a more extreme manner, throughout the last two centuries localizationist approaches have been countered by holists or advocates of equipotentiality such as Flourens and Lashley. Some advocates of dynamical systems theory today espouse a similar holism and repudiate any attempt to isolate distinct neural systems responsible for specific tasks. See, for example, van Orden, in press. Emphasis on functionally and anatomically distinct systems is, however, the rule rather than the exception in cognitive science.

Horgan and Tienson (1994, pp. 319) consider the possibility that human cognition might be best modeled through utilization of a number of distinct, interacting sub-systems. They argue, however, that from a dynamical systems perspective ‘coupled dynamical systems’ can always be analyzed mathematically as a single, “larger, higher-dimensional, total dynamical system . . .” (pp. 319). It may be, however, that analyzing functionally distinct, coupled dynamical systems in terms of a total dynamical system will both necessitate the rejection of the quest for tractable computation over representations and, at the same time, seriously undermine ongoing research which is best understood in terms of such computation over representations. An example should make these points clear.

Research into speech perception has suggested the existence of a number of distinct, interacting subsystems devoted to solving selected aspects of the task of recognizing spoken words (see Best 1992 for an overview). Early encoding of speech is in the form of a representation of raw acoustic information. This information is subsequently analyzed for certain distinctive features corresponding to the manner in which the speech was produced. Production features include, for example, such properties as voice onset time (the degree of synchrony between the release of air and the onset of vocal chord vibration) and place and manner of articulation (the site and degree, respectively of closure of the vocal cavity). At an algorithmic level of description, we may say that the cognitive system is equipped with detectors which are sensitive to these production features (see Liberman et al. 1967).[5] Individual speech sounds (phonetic segments) are subsequently identified through unique combinations of these features (Clark and Clark 1977). Speech sounds, in turn, are further processed by systems responsible for lexical recognition and are simultaneously read into a temporary phonological store (see Martin and Breedin 1992). In addition, there seem to be important top-down influences on the recognition process. These include influences of implicit knowledge of the manner in which the production of a given sound will influence the production of subsequent sounds (Pollack and Pickett 1964), and semantic expectations generated by sentence context (Warren and Warren 1970).

Interestingly, there are a number of properties of the speech recognition system which suggest that a connectionist implementation would be the most natural way to model the speech recognition process. For instance, speech recognition seems to be accomplished through multiple levels of pattern matching; the system is able to handle a great deal of variability and noise; it seems to learn through early exposure to the phonemes characteristic of one’s native tongue (see Mazoyer et al. 1993); and there is both bottom-up and top-down processing as well as sensitivity to temporal constraints.

With regard to the present debate, this example illustrates the manner in which cognition is often understood in cognitive science. A realistic connectionist model of a complex cognitive task would consist of a number of distinct, though tightly coupled, networks (for an example, see Miikkulainen, 1993). One could, of course, analyze the system as a whole in terms of a complex activation landscape. If one chooses to approach the task this way, it may seem that the processing will be intractable and only a dynamical systems analysis will be informative. However, there may be good reasons why natural systems exhibit a more modular structure, one in which functions are decomposed and localized in modules (Bechtel and Richardson, 1993)--such systems may be both

[5]There seems to be promise that such detectors might be identified at the physical (viz. neuronal) level. Eimas and Corbit (19--), for instance, have shown that these feature detectors can be fatigued in a manner that is strikingly similar to the fatiguing of visual feature detectors.

more likely to evolve and more efficient. The argument from evolution stems from Herbert Simon (1980), who advances the analogy of two watchmakers to show the advantages of a modular system. Each makes an equally fine watch of 1000 different parts. One watchmaker, Tempus, uses a modular design in which each group of 10 components forms a stable subsystem and in which groups of ten subsystems again form stable structures. The other watchmaker, Hora, employs a design in which all 1,000 pieces must be correctly aligned before the whole becomes stable. Every time Hora answers the phone to take an order, all of his work is lost; Tempus, on the other hand, loses only the last segment of work. From an evolutionary perspective, the point is that if pieces could evolve separately, then evolving a complex system such as a human would not take such an improbable event as all the parts coming together at once. A further evolutionary advantage is that one could modify one of the parts without risking loosing the proper performance of all other functions.

The argument from efficiency is illustrated in recent connectionist modeling in the related domain of visual processing. Mishkin, Ungerleider, and Macko (1983) presented evidence of two different pathways in the visual system, one of which processes information about the spatial location of objects while the other processes information about the identity of objects. In an attempt to understand why the visual system might process *what* and *where* information separately, Rueckl, Cave, and Kosslyn (1989) developed a connectionist simulation in which the network was trained to identify an input pattern on one set of units and specify its location on the input grid on another set of units. They showed that the network learned the task more rapidly if two distinct sets of hidden units were employed, one feeding the *where* output units, the other feeding the *what* output units. Jacobs, Jordan, and Barto (1991) explored this issue further and showed that a modular network could learn to decompose the two tasks for itself even when it used only one set of output units.

In the simulation by Jacobs et al., the input consisted of an activation pattern on a 5x5 grid and a unit specifying either the object or its location (see Figure 2). The system contained three *expert networks*, each with different designs: one was a two-layer network, and two were three layer networks with either 18 or 36 hidden units. In addition to these, the system contained a gating network, which would determine by its output how much influence each of the three expert networks would have in determining the answer to a specific query. The individual expert networks were trained by a variation of backpropagation in which the strength of the error signal was proportional to the degree of influence the expert network had on determining the overall answer. The gating network was trained through a somewhat more complicated procedure which increased or decreased the influence of an expert network on a particular kind of problem (*what* or *where*) depending on whether that expert network was more or less accurate than the others.

The modular system learned the task with 50 epochs of training, and divided the *what* and *where* tasks to different expert networks: the *where* task was performed by the two-layer network while the *what* task was performed by the three layer network with 36 hidden units. The authors' explanation is that the *where* task is linearly separable, and hence one that could be performed quickly and accurately by the two-layer network, whereas the *what* task was not linearly separable and so required the use of hidden units (see Bechtel and Abrahamsen, 1991 for an exposition of linear separability). They compared this performance with that of a single three-layer network trained on both tasks. After 100 epochs, that network achieved only 80% accuracy. Moreover, it exhibited catastrophic interference: if the *what* problems were presented as a block, followed by the *where* problems as a block, and it was then tested on the *what* problems, performance would be compromised. The modular system showed no such interference.

Connectionist simulations with modular network are thus highly suggestive as to why the mind/brain might decompose tasks into component subsystems. In addition to better learning and avoiding catastrophic interference, Jacobs et al. argue that modular networks will likely facilitate more reasonable generalization (because distinct parts of the task will be performed by networks that each generalize well for their tasks), more intelligible and useful representations (because each network develops representations for particular tasks), and more efficient use of computational hardware (since each module will only have to represent a more limited set of dimensions). Thus, a componential system, with modules performing tractable computations over representations may prove to be better suited to modeling cognitive tasks than large, non-modularized system. Moreover, we suggest that it has been by seeking such systems that cognitive scientists have made the most inroads to understanding human cognition.

2. Need Representations be Syntactically Structured?

As Horgan makes clear, one of the major claims made on behalf of connectionist networks is that they do not employ syntactically structured representations in the manner of classical systems. For some critics, such as Fodor and Pylyshyn (1988), this has represented a shortcoming of connectionist models. Without syntactically structured representations, they claimed that

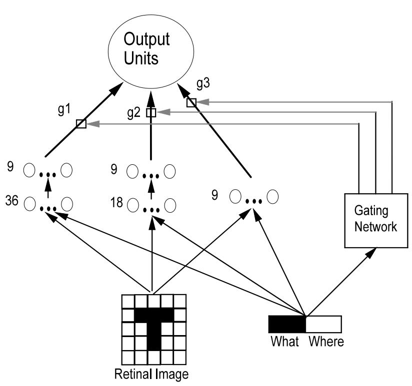

Figure 2: The network used by Jacobs et al. (1991) to simulate the *what/where* task. The inputs specify which units in the retinal image are active and whether the task is to identify the pattern or its location. This whole input is provided to the three expert networks, but only the task specification is provided to the gating network. All three expert networks process the input, but the gating network determines how much each of them will influence the pattern produced on the output units.

connectionists would be unable to explain such features as the productivity and systematicity of thought. There have been a number of connectionist responses to this objection. One is to develop connectionist systems that employ another binding technique that does the same work as syntactic structure (such as synchrony, see Shastri and Ajjanagadde, 1993). Another is to argue that productivity and systematicity only emerge to the degree Fodor and Pylyshyn claim they do with language using creatures and that it is the ability to use natural language representations that gives humans the degree of productivity and systematicity that is observed (Clark, 1989, Dennett, 1991, Bechtel, 1996). What is different about Horgan's proposal is that he proposes to answer Fodor by keeping syntactically structured representations in connectionist models.

Although Horgan defends the use of syntactically structured representations, his portrayal of them differs significantly from that endorsed in classicism. The difference is that they are not operated on by rules that respect their syntactic structure. To determine exactly what the preservation of syntax comes to for Horgan, we need to examine his proposal in some detail. He adopts a dynamical system's perspective on connectionist networks and proposes that we represent a network in a multidimensional space in which each unit defines a dimension and in which the pattern of activation in a network at a time is mapped to a point in that space. Processing in the network will result in a trajectory of this point through that space. Since complex cognitive states, such as those involving syntactically structured representations, are to be rendered at this level as points, the relation between the cognitive characterization of the state and the dynamical system characterization of the state cannot be one of *intrinsic* realization in which the structure of the cognitive state mirrors that of the representational state, since the representational state is a point and so has no structure. Rather, the relation is a *dispositional* one in which the realization (the point in the multidimensional space) preserves the counterfactual profile of the cognitive state. (The distinction between intrinsic and dispositional realization is found in Horgan and Tienson, 1994.) This means that any possible transitions between cognitive states (e.g., in the course of thinking or learning) will mirror possible trajectories in the multidimensional space.

Before we can evaluate what syntactically structured representations amount to for Horgan, we need to sketch some background. In defending the classical account, Fodor and Pylyshyn (1988) argued that an important feature of the classical perspective was the use of a compositional syntax in which parts of a complex representation stood for component elements and the relations between the parts of the representations stood for relations between elements. van Gelder (1990), in arguing that connectionists provided a different conception of representation, proposed that while classical systems insisted on concatenation as the mode of compositionality, connectionist modelers such as Smolensky (1990) and Pollack (1990) provided functional compositionality. His point was that their networks preserved information about syntactical structure without using concatenation. Specifically, one could not identify in the representations constructed in these networks parts corresponding to the elements of what was represented.

Horgan and Tienson (1994) cite both of these examples; since Pollack's RAAM (Recursive Auto Associative Memory) networks are easier to present, we will focus on it here. Pollack's overall network shown in Figure 3a is an autoassociative network, which means that it is trained to reproduce on its output units the same pattern as is presented on its input units. The units in the hidden layer constitute a compressed representation. The network is recursive in that, if one is encoding something that has a tree structure, such as the one shown in Figure 4 (any list of items, such as a sentence, can be construed as a tree structure), one first forms compressed representations of the items on the terminal nodes, and uses the compressed representations that are formed for these items as one moves up the tree. When one reaches the top of the tree, one has a compressed representation of the whole tree. Once the network has been trained, one can separate the two parts of the network, using the bottom part (Figure 3b) to encode the tree and the top part (Figure 3c) to decode it. The interest in this procedure is that you can use the compressed representations as you would use the whole syntactically structured sentence in other networks to, for example, construct

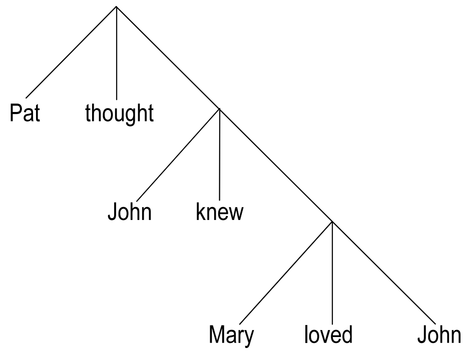

Figure 4: A tree representation of the sentence 'Pat thought John knew Mary loved John.'

the passive (Chalmers, 1990) or to answer queries as to whether specific words or phrases were present in the original sentence (Blank, Meeden, and Marshall, 1992).

We can now see what syntactically structured representations amount to for Horgan. As we have already noted, they do not have parts corresponding to elements of what was represented. Rather, the point corresponding to the activation pattern in the compressed representation is related to other points in a way that preserves the relations between the elements that would be expressed in terms of concatenational syntax in language. In the case of RAAM, as the network is learning to reconstruct the same patterns (corresponding to sentences) on the output as provided on the input, it gradually changes the pattern of activation constructed on the hidden units (and so it changes the point it occupies in the multidimensional space as a result of a given input). Horgan and Tienson describe this as follows: “This repositioning reflects a refinement of the realization relation from intentional states to points on the activation landscape. The realization relation exhibits increasing systematicity, coming to reflect, in the way it positions representation-realizing points relative to one

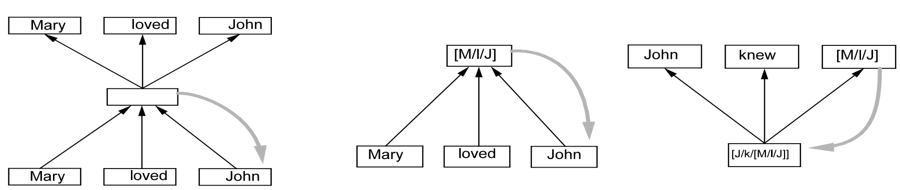

Figure 3: The RAAM network developed by Pollack (1990) to encode compressed representations of the tree structure shown in Figure 4. (a) The full autoassociative network used during training. Here the terminal nodes from the last right hand branch of the tree are encoded on the input and output units, and the network is trained to regenerate on the output units the same pattern as is presented on the input units. The compressed representation that is formed is then copied back onto the rightmost set of input units as the network progresses up the tree. (b) The lower half of the network shown in a, which is used after training as the encoding network. (c) The upper half of the network shown in a, which is used after training as the decoding network.

another on the activation landscape, important relations among the intentional states being realized" (ms. p. 25).

RAAM networks appear to provide an interesting way to represent syntactically structured information in a manner that is only functionally, not explicitly, compositional. But it is important to realize why a RAAM's compressed representations preserve functionally syntactic information. The reason has to do with the task that is being performed. The network is trained to reconstruct from the explicit encoding of syntactically structured sentences outputs that exhibit that same explicit encoding. It is in fact essential to the operation of a RAAM network that it not be asked to reconstruct all possible input patterns, but only those which are structured in some way. Otherwise, the task will exceed the informational capacity of the compressed representation. In Pollack's work, he employed several structured inputs: either letter sequences corresponding to English words, strings of grammatical categories permitted by a grammar, or actual sentences of English (the example used in Figures 3 and 4 comes from this last experiment). Once trained, his networks were able to generalize to new cases, but ones that employed the same structuring principles as the training cases. Thus, it was when the network was trained on syntactically structured inputs that it developed compressed representations that preserved syntactical structure. The network did not create a system of syntactically structured representations *de novo*, it transformed a structure it was given into another format, one that used less resources.

Horgan and Tienson are certainly right that one of the things a network acquires as it learns are dispositions to respond to inputs which preserve the transformations specified in the training set. But the question to be raised is whether these dispositions will really correspond to transformations that respect syntactical structure. We will argue that (a) except possibly for networks trained to perform language tasks, once one gives up classicism at the algorithmic level, there is no reason to retain syntactically structured representations, (b) while tasks involving language are the most natural ones to model using syntactically structured representations, they may not be necessary even for modeling language tasks, and (c) a more natural ally for connectionist modeling is cognitive grammar, which construes syntactic structure as derivative from more fundamental cognitive capacities, which rely on semantic structures rather than syntactic ones. Thus, we will argue that connectionists were indeed right to reject syntactically structured representations.

a. Why retain syntactically structured representations outside of language tasks?

One thing that is distinctive about Pollack's and Smolensky's networks is that they are networks designed explicitly to process syntactically structured representations, specifically, syntactically structured representations of natural language. If one's goal is to carry out in a network operations that are defined upon syntactically structured representations, then it could seem reasonable to build up representations in a network that at least functionally maintain the information about syntactic structure. But the point to be emphasized is that it is the task that one has set for the network that imposes this constraint. The question to ask, then, is whether the broad range of cognitive tasks an agent confronts, such as navigating an environment or physically acting in it, imposes similar demands. When training a network to handle these tasks, will it develop transitions between internal states such that we will identify these states as encoding syntactic structures? The answer would seem rather clearly to be "no". The network will develop some transition dispositions, but these will be geared to the task at hand.

This point is really just an extension of the opening up of a broader range of mathematical functions for modeling cognitive processes that Horgan emphasizes as a virtue of connectionism. The reason for using syntactically structured representations in classical systems is that the very operations such systems perform operate on the structure of those representations. The structure of the representations and the operations must be fitted to each other. It is likewise true that the structure of the representations used in connectionist networks must be fitted to the operations performed on the representations. But these operations are not limited to ones that operate on the syntactical structure of representations. Moreover, to get the network to behave in ways that fit the syntactic structure of representations takes special training (especially since that structure is not there in the representations themselves but only implicitly there as a result of the way in which the network learns to respond to a pattern of activation). But, and this is the important point, the network can be trained to respond to activation patterns in various different ways, not just ones specified in syntactic rules. Since we are to read the structure of the representation off of the transition dispositions, this structure will not be syntactic one unless the transition dispositions adhere to that syntax.

b. Are syntactically structured representations even required for language tasks?

The processing task for language using creatures, whether they use language primarily for communication or in the course of their own private thinking, is to encode information into linguistic structures and extract information from them. It may seem that the best way to develop a system to work with such structures is first to build up an internal representation of those structures and then to operate on those internal representations. But in a variety of domains cognitive scientists are coming to realize that we can operate on information coming in from the world without building up a complete representation internally. Rodney Brooks (1991) has argued that the world is its own best representation. This does not mean, as Brooks sometimes suggests, that the cognitive system does not need to represent features of the world. It does need representations, although possessing representations only involves having states that carry information about possible features of the world which are used by the system in order to coordinate its behavior with the entities in the world (Bechtel, in press). When the entities in the world are linguistic structures, it needs to acquire, and so represent, features of these linguistic structures.

The question then is whether a cognitive system can engage in behavior employing language with less than a comprehensive internal representation of linguistic structures; in particular, can it function with representations that are not compositional, functionally or structurally? Until there are networks which exhibit the ability to communicate and to reason to themselves in natural language without building up syntactically structured internal representations, we will not be able to claim definitively that this is possible. But there is a variety of research using networks in language related tasks which suggests that it is possible. Many of these simulations use recurrent networks (Figure 5), in which patterns of activation produced on hidden units in one cycle of processing are fed back in as part of the input at the next cycle (e.g., next word), either to process linguistic strings one word at a time, or to produce linguistic strings one word at a time. Some of these networks build up case role or schema based representations of the meaning of sentences or stories (St. John and McClelland, 1990; Miikkulainen, 1993). To illustrate the basic idea that a network might encode specific information that it needs in order to operate with natural language structures without building up an explicitly or functionally compositional internal representation, we will describe a relatively simple task employed by Elman (1990, 1991; see also Christiansen, 1994).

In Elman's research, networks are trained to predict the next word in a linguistic corpus. The corpus is constructed from a grammar, which distinguishes transitive and intransitive verbs and allows multiply embedded relative clauses. (Christiansen's simulations use a significantly more complex grammar which, in addition, allows multiply embedded genitives, prepositional phrases, and propositional attitude structures.) Since the corpus of text is relatively large (10,000 sentences during each epoch of training), and the network was only trained with a few passes through the corpus, it could not identify the single next word in the utterance; rather, it would respond with partial activations to numerous words, and it was evaluated by whether these words belonged to appropriate grammatical categories for possible continuations of the sentence. The challenge for the network was to identify only words that would respect the grammatical principles, for example, selecting plural verbs after a plural subject, even if one or more relative clauses intervened. Elman showed that with an appropriate training regime[6], networks could learn to make highly accurate predictions. (In Christiansen's simulations with more complex grammatical structures, more errors appeared, but he showed that many of these correspond to the types of errors humans exhibit in natural language use.)

The question arises as to how the network achieves this kind of performance. It does not have a traditional memory structure in which to build up an explicit representation of whole sentences. Recurrent connections allow the network only to keep a constantly degrading trace of some aspects of previous cycles (which aspects it retrains depends upon the task the network is

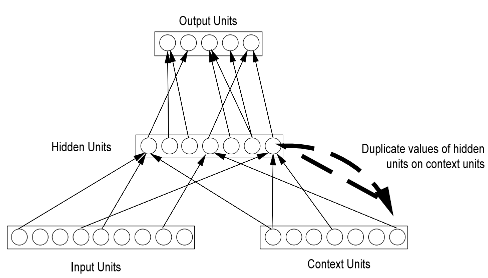

Figure 5: A simple recurrent network designed by Elman (1990). As each new input is supplied to the input units, the activation pattern from the previous cycle is copied over onto the context units and is treated as part of the input. This provides the network with a constantly degraded trace of previous cycles of processes. This turns out to be sufficient for networks to learn complex sequential structures specified in grammars allowing for recursively embedded clauses.

required to learn). Given that Elman's network employed 70 hidden units, there is no simple way to determine what information the network retained and how it did so (that is, one cannot simply

---

[6]The procedures were either one that used only simple sentences in the first epochs of training, or one which reset the recurrent pathways after a limited number of processing cycles during the early epochs of training. Without this technique, the network would not first learn the elementary, subject-predicate structures of the language. Elissa Newport (1988, 1990), after discovering that late learners of a language (e.g., learners of a second language after childhood) never reached the level of performance of early learners, suggested that children have an advantage in that they bring less cognitive processing ability to the task, which causes them to master basic structures first. She dubbed this the less is more hypothesis.

evaluate the contributions of each hidden unit). Elman has, therefore, used a variety of other techniques to analyze how the network behaves, such as cluster analysis and principal components analysis. Cluster analysis groups together input patterns in terms of the hidden unit patterns they generate; Elman's (1990) cluster analysis of a simpler network revealed that it grouped words together by their grammatical categories so that, for example, verbs used only transitively were grouped together, and included in the broader category of verbs. With his more complicated network, Elman (1991) instead employed principal components analysis, which provides a lower dimensional analysis of hidden unit activations which can then be analyzed to see how information needed for a specific function (e.g., a specific subject-verb agreement) is maintained in the hidden units. What this analysis suggests is that his network is retaining information that it will later need (by producing variant patterns of hidden unit activations) until the information is no longer needed, at which point the network no longer retains it.

There is no reason to think that the representations in Elman's networks have a compositional structure, explicit or functional. Rather, they seem to retain various pieces of information about previous input patterns by producing somewhat different hidden unit activations. This suggests a very different way of thinking about the task of networks, whether they use natural language or not: what they acquire is procedural knowledge and what they represent is information that is needed to perform specific procedures. Even when the items with which they are working (in this case, word sequences) are syntactically structured, they do not need to build up a complete representation of that structure. They may need to represent some information about that structure (for example, that a subject for which a verb has not yet been encountered is singular), but this is done on a need-to-know basis. In the case of language, then, what is represented (actual discourse) will be syntactically structured, but the mental or network representation need not be.

c. Cognitive Linguistics: Syntax and Connectionist Modeling.

Most theorizing in cognitive science about language has adopted the framework of generative grammar. A guiding tenet of the generative grammar tradition is a strong distinction between syntax and semantics and the proposal that an exhaustive understanding of grammar can be obtained by considering syntax in isolation (what Langacker, 1991a, p. 515, a critic, terms the 'autonomy thesis'). A further aspect of this approach has been the assumption that human knowledge of syntax would take the form of a grammar that could generate all and only the grammatically well-formed sentences of a language. Chomsky's early work on grammar demonstrated that grammars adequate to natural languages required a computational system of the same power as a universal Turing machine (Chomsky, 1957), which seems to support the claim that the human cognitive system must itself have the power of a universal Turing Machine. Subsequently Chomsky differentiated between competence and performance, allowing that in performance memory limitations and other factors might account for the production of ungrammatical sentences. He retained the idea, though, that the internal processing system is a powerful computational device comparable to a Turing machine. However, the approach we presented in the previous section suggests a very different outlook (see also Bechtel, 1996). Human cognitive activities may be able to interact with natural language representations (which are themselves syntactically structured) without requiring these symbols to be inside the system (that is, without requiring the tape of the Turing machine to be in the head). The approach we now consider takes this one step further by questioning whether we should accept the autonomy thesis; as we shall see, denying it opens the way to a very different account of what is required to comprehend and produce syntactically well-formed sentences.

In marked contrast to the generative grammar tradition stands cognitive linguistics, a relatively new approach to linguistics with its own set of constraints on linguistic explanation. One constraint, which provides the ubiquitous and unifying thread of cognitive linguistics, is what Langacker has termed 'the content requirement' (1990, p. 18-19). The content requirement expands the scope of linguistic explanation to include the semantic domain and treats knowledge of language as consisting in knowledge of correspondence relations between (and constraints on integration of) linguistic and semantic structures. More specifically, cognitive linguistics countenances only two kinds of structures (sometimes referred to as "poles") and relations between them: (a) linguistic items which are overtly tokened in public or private speech (these vary in level of complexity from morphemes to clause-level structures and are termed 'phonological units'), and (b) semantic structures (these vary in level of complexity from specific to schematic and are termed 'semantic units') (p. 16-19). To begin with word meanings, knowledge of a word such as "pencil" consists at one pole of phonological knowledge of the word and at the other pole semantic knowledge (which, for example, enables one to recognize or label pencils in various contexts) and relations between these poles.

What is pertinent for our purposes is that knowledge of syntax is handled in the same manner. Specific morpho-syntactic structures (e.g. the suffix '-ed', the passive morphology 'be *verb*-ed by', modal verbs, and so forth) are linked with semantic knowledge. Each syntactic structure has a semantic import. Evidence for this comes from considering pairs of sentences such as the following which both employ the same syntactic form, but where the semantic pole for one is incoherent:

(1) Cynthia threw Fred the ball.

(2) *Cynthia threw the wall a dart.

(1) exemplifies the English di-transitive construction which follows the syntactic form [Subj V Obj_1 Obj_2] (Goldberg 1995, ch. 7). Cognitive linguists proposes that this syntactic form is linked to the schematic semantic structure [X causes Y to receive Z]. This schematic semantic structure constrains the type of units which may properly act as constituents of the di-transitive construction. One constraint is that the entity named by Obj_1 must be capable of functioning as a recipient. In (1) there is an entity (Fred) symbolized by Obj_1 which can fill this role and thus be integrated into the clause level semantics. In (2), however, this semantic constraint is not satisfied. "The wall" does not designate something which can count as a recipient in this context, thus preventing a coherent integration of the clause-level semantics with the semantics of this constituent. It is this failure of integration (and not the failure to accord with knowledge of proper syntactic form) which explains, according to cognitive linguists, the perceived ungrammaticality of sentence (2).

Connectionism and cognitive linguistics have been widely recognized as natural allies (Harris 1989, Langacker 1991a, Regier 1995, Goldberg 1995), and we can already see part of the basis for this. The mappings between the two poles that cognitive linguistics count as constituting the basic knowledge of language is something naturally modeled in networks. Langacker is explicit in making the connection:

> the acquisition of a grammar (as of lexicon) becomes a matter of establishing form-meaning mappings, which PDP systems handle straightforwardly as associations between input and output patterns (Langacker 1991a, p. 533).

This is particularly natural insofar as cognitive linguists construe grammatical knowledge as being embodied in cognitive systems in the form of *implicit* or procedural knowledge (Langacker 1991a, p. 533-535, see also Tomasello, forthcoming). They also view the mapping process as occurring through the satisfaction of multiple constraints (Langacker 1991a, Harris 1989, see also McClelland and Kawamoto 1988), something naturally accomplished in networks. In addition, there are features of connectionist learning that can implement aspects of language emphasized by cognitive linguists. For instance, the need for the extraction of prototypes exhibiting graded structure through positive exemplars, the ability to generalize to novel cases based upon similarity to known cases, and the ability to handle regularities and subregularities are all naturally accommodated through connectionist learning procedures (Harris 1989, Regier 1995).

There is a further respect in which cognitive linguistics helps show how linguistic knowledge can be modeled without requiring syntactically structured mental representations. This is best seen if we consider how cognitive linguists propose to model more complex linguistic structures, such as embedding one syntactic structure in another. Langacker refers to simple syntactic structures such as the di-transitive construction as "units" and explains:

> The term "unit" is employed in a technical sense to indicate a thoroughly mastered structure, i.e. one that a speaker can activate as a preassembled whole without attending to the specifics of its internal composition. A unit can therefore be regarded as a cognitive routine. The inventory of conventional units is "structured" in the sense that some units function as components of others (i.e. they constitute subroutines) (p. 15).

This is illustrated in

(3) Cynthia passed Tom Irene's cap.

In (3) the di-transitive construction is integrated with the possessive construction. At the phonological pole subordinate structures are combined to form more complex ones (e.g. through the 'Subj Verb Obj Obj' and '-'s' morphology), and the same occurs at the semantic pole. Lexico-syntactic morphology, furthermore, is taken to be a reflection of conceptual structuring (Langacker 1991b, p. 24). Importantly, however, while the manner of integration at the phonological pole is sometimes concatenative in nature,[7] the result of integration at the semantic pole is an holistic conceptualization (often conceived by analogy to the overlaying of transparencies). Since the phonological pole is only realized in the public (or privately rehearsed) sentence, there is no assumption of an underlying mental concatenated structure lying behind the sentence itself. Rather, the only underlying mental structure is the holistic semantic structure. If cognitive linguists are able to account for the range of syntactic structures found in natural languages in terms of underlying semantic structures, then cognitive linguistics will provide further support for our argument in section (b) that even when networks are engaged in linguistic tasks, they do not need to employ syntactically structured representations. Knowledge of syntax may be understood as being embodied by a system (or systems) in the form of procedural knowledge of the constraints governing the mappings between (in Langacker's terminology) the semantic and phonological poles.

3. Conclusion

Horgan argues that the strongest case for connectionism as a basis for cognitive modeling involves construing it as rejecting "tractably computable cognitive transition functions" but yet employing syntatictically structured representations. We have argued against both of these claims. With respect to tractable computation, we pointed out that it is not clear that forgoing tractable computability will solve the problem which Horgan seeks to solve, explaining the Quinean and isotropic nature of cognition. We also argued that human cognition is not nearly as Quinean and isotropic and Horgan proposes. Most importantly, we tried to show how progress in cognitive science is currently be made by those seeking tractable computational explanations, albeit not at the

[7]Though integration at the phonological pole is often concatenative, the set of licensed expressions is ill-defined and hence the possibility of a generative algorithm which outputs all and only grammatical expressions is precluded (see Langacker 1991b, 15-19).

level of the whole system. But we argued that is not the level at which cognitive science should aspire to operate. With respect to syntactically structured representations, we argued that for most cognitive tasks syntactically structured representations are unnecessary. Further, we argue that there is good reason to believe that even linguistic tasks can be modeled without linguistically structured representations. We finished our argument by noting that cognitive linguistics, not generative linguistics, is the natural ally of connectionism, and that the program of cognitive linguistics, by abandoning the assumption of autonomous syntax, suggests that knowledge of language does not require syntactically structured internal representations. In the course of advancing these arguments we have also presented several examples of recent connectionist research which we hope provides a good introduction to the current developments in this domain.

References

Bechtel, W. (1996). What knowledge must be in the head in order to acquire language. In B. Velichkovsky and D. M. Rumbaugh, (Eds.), *Communicating meaning: The evolution and development of langauge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bechtel, W. (in press). Yet another scientific revolution? Defusing the dynamical systems theorists' attack on mental representation.

Bechtel, W. and Abrahamsen, A. A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Oxford: Basil Blackwell.

Bechtel, W. and Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Beer, R. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*, 173-215.

Blank, D. S., Meeden, L. & Marshall, J. B. (1992). Exploring the symbolic/subsymbolic continuum: A case study of RAAM. In J. Dinsmore (ed.), *Closing the gap: Symbolism vs. connectionism*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47, 139-159.*

Chalmers, D. J. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, *46*, 47-75.

Chomsky, N. (1957) . *Syntactic structures*. The Hague: Mouton.

Christiansen, M. H. (1994). Infinite languages, finite minds. Connectionism, learning, and linguistic structure. Ph.D. Dissertation, University of Edinburgh.

Clark, A. (1989) *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press/Bradford Books.

Clark, A. (1993). *Associative engines: Connectionism, concepts, and representational change.* Cambridge, MA: MIT Press.

Clark, A. and Karmiloff-Smith (1994). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language*, *8*, 487-519.

Dennett, D. C. (1991). *Consciousness explained.* Boston: Brown Little.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195-225.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

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

Haselager, W. F. G. (1995). *The right frame of mind: Cognitive science, folk psychology, and the frame problem.* Ph.D. Dissertation, Vrije Universiteit Amsterdam.

Hinton, G. E. and Anderson, J. A. (eds) (1981) *Parallel Models of Associative Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Horgan, T. and Tienson, J. (1991). *Connectionism and the philosophy of mind*. Dordrecht: Kluwer.

Horgan, T. and Tienson, J. (1994). A nonclassical framework for cognitive science. *Synthese*, *101*, 305-45.

Horgan, T. and Tienson, J. (1996). *Connectionism and the philosophy of psychology*. Cambridge, MA: MIT Press.

Jacobs, R. A., Jordan, M. I., and Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, *15*, 219-250.

Kelso, J. A. S. (1995). *Dynamic patterns: The self organization of brain and behavior*. Cambridge, MA: MIT Press.

McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie (Eds.), *Machine Intelligence*, pp. 463-502. Edinburgh: Edinburgh University Press.

McClelland, J. L., Rumelhart, D. E., and the PDP Research Group (1986) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, MA: MIT Press/Bradford Books.

McCloskey, M. and Cohen, N. J. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*, vol. 24, pp. 109-65. New York: Academic Press.

Miikkulainen, R. (1993). *Subsymbolic natural language processing. An integrated model of scripts, lexicon, and memory.* Cambridge, MA: MIT Press.

Minsky, M. A. and Papert, S. (1969) *Perceptrons*. Cambridge, MA: MIT Press.

Mishkin, M. Ungerleider, L. G., and Macko, K. A. (1983). Object recognition and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*, 414-417.

Newell, A. (1989). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newport, E. (1988). Constraints on learning and their role in language acquisition: Studies in the acquisition of American Sign Language. *Language Studies*, *10*, 147-172

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11-28.

Pollack, J. (1990). Recursive distributed representations. *Artificial Intelligence*, *46*, 77-105.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.) *Semantic Information Processing*, pp. 227-70. Cambridge, MA: MIT Press

Ramsey, W., Stich, S. P., and Rumelhart, D. E. (1991). *Philosophy and connectionist theory*. Hillsdale, NJ: Lawrence Erlbaum.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285-308.

Rosenblatt, F. (1962) *The Principles of Neurodynamics*. New York: Spartan.

Rosenbloom, P. S., Laird, J. E., Newell, A., and McCarl, R. (1991). A preliminary analysis of the SOAR architecture as a basis for general intelligence. *Artificial Intelligence*, *47*, 289-325.

Rueckl, J. G., Cave, K. R., and Kosslyn, S. M. (1989). Why are 'what' and 'where' processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, *1*, 171-186.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning representations by back propagating errors. *Nature*, *323*, 533-6.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1986) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books.

St. John, M. F. & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217-257.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*, 159-216.

Simon, H. A. (1980). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Thelen, E. (1995). Time-scale dynamics and the development of an embodied cognition. In R. Port and T. van Gelder (eds.), *Mind as motion*. Cambridge, MA: MIT Press.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, *14*, 355-384.

van Orden, C. G. (in press). What do double dissociations prove? Inductive methods and theory in psychology.