A Connectionist Architecture with Inherent Systematicity

James Henderson
Department of Computer Science
University of Exeter
Exeter EX4 4PT
United Kingdom
jamie@dcs.ex.ac.uk

Abstract

For connectionist networks to be adequate for higher level cognitive activities such as natural language interpretation, they have to generalize in a way that is appropriate given the regularities of the domain. Fodor and Pylyshyn (1988) identified an important pattern of regularities in such domains, which they called systematicity. Several attempts have been made to show that connectionist networks can generalize in accordance with these regularities, but not to the satisfaction of the critics. To address this challenge, this paper starts by establishing the implications of systematicity for connectionist solutions to the variable binding problem. Based on the work of Hadley (1994a), we argue that the network must generalize information it learns in one variable binding to other variable bindings. We then show that temporal synchrony variable binding (Shastri and Ajjanagadde, 1993) inherently generalizes in this way. Thereby we show that temporal synchrony variable binding is a connectionist architecture that accounts for systematicity. This is an important step in showing that connectionism can be an adequate architecture for higher level cognition.

Introduction

Connectionist networks have been successfully applied to a wide variety of problems, but they have not had much success with higher level cognitive activities. For example in natural language parsing, connectionist networks have not been able to exhibit the same generalization abilities that statistical approaches based on standard parsing techniques have shown. This inability to generalize in the appropriate way can be attributed to an inadequacy in standard connectionist networks that was first identified by Fodor and Pylyshyn (1988). They pointed out a pattern of regularities in higher level cognitive activities that they called systematicity, and challenged connectionists to show how this pattern could be accounted for within connectionism. In this paper we show that a particular connectionist architecture, called temporal synchrony variable binding (Shastri and Ajjanagadde, 1993), accounts for systematicity through its inherent ability to generalize across entities. In the process, we clarify the notion of systematicity and show its implications for variable binding techniques. This work not only contributes to the continuing debate about systematicity in connectionist networks, it also indicates how connectionist networks' impressive learning abilities for pattern matching tasks can be extended to the more complex domains typical of higher level cognitive activities.

As Fodor and Pylyshyn (1988) describe it, systematicity embodies the kinds of regularities that arise from a compositional generative grammar. Because in such grammars a general-purpose composition operation is used to construct sentences out of their constituents, a given constituent can appear anywhere that its type of constituent is allowed. For example, because "John" and "Mary" are the same type of constituent, a generative grammar would not generate "John loves Mary" without also generating "Mary loves John". Fodor and Pylyshyn (1988) argued that such regularities are an inherent part of human cognitive activity, and thus that connectionism could not be an adequate cognitive architecture until it accounted for this phenomena. Many attempts have been made to meet this challenge (e.g. Smolensky, 1990; Christiansen and Chater, 1994; Niklasson and van Gelder, 1994), but the critics have not been satisfied (Fodor and McLaughlin, 1990; Hadley, 1994a; Hadley, 1994b).[1]

Such difficulties have lead some connectionists investigating higher level cognitive activities to propose extensions to standard connectionist architectures. One such investigation developed a technique called temporal synchrony variable binding for use in fast, common sense reasoning (reflexive reasoning) (Shastri and Ajjanagadde, 1993).[2] This technique can represent multiple features of multiple entities, and it can perform a significant class of computations over this representation. In addition to reflexive reasoning, it has been successfully applied to syntactic natural language parsing (Henderson, 1994b; Henderson, 1994a). This work showed that the kinds of regularities that systematicity embodies could be directly and simply represented in a network that uses temporal synchrony variable binding. In this paper we show that it is not only possible to build a connectionist network that exhibits these regularities, but that these regularities are a consequence of the inherent generalization abilities of temporal synchrony variable binding networks.

Because there has not been general agreement about the

[1] One proposal (Hadley and Hayward, 1995) has not yet been criticized in the literature. This work will be discussed below in the section on representing variable bindings.

[2] The possibility of using temporal synchrony for encoding feature bindings in the perceptual domain was suggested by von der Malsburg and Schneider (1986), but Shastri and Ajjanagadde were the first to use temporal synchrony to solve complex representational problems in higher level cognitive activities.

exact definition of systematicity, this paper starts with a discussion of this issue based on the work of Hadley (1994a). The resulting definition of systematicity requires a solution to the variable binding problem where learned parameters are independent of variable bindings. Several connectionist approaches to variable binding are then discussed with respect to this requirement. For most of the proposed methods, there is no apparent way for them to satisfy this requirement. In contrast, temporal synchrony variable binding by its very nature makes learned parameters independent of variable bindings. By showing this we show that temporal synchrony variable binding is a connectionist architecture that can account for systematicity.

The Definition of Systematicity

Although neither (Fodor and Pylyshyn, 1988) nor (Fodor and McLaughlin, 1990) provides a precise definition of systematicity, the concept is meant to embody the pervasive regularities in language (and thought) that are traditionally captured using compositional generative grammars. Such grammars express generalizations about the sentences of a language in terms of their subconstituents. A universal composition operation is then used to combine these constituents to form sentences. Because generalizations are stated in terms of constituents, such an architecture predicts a prevalence of regularities across constituents. This is in fact the case. For example, the pattern of words that can make up a noun phrase is very complex, yet this pattern is virtually identical for subject and object noun phrases. Thus if "John" can be a subject noun phrase and "Mary" can be an object noun phrase (as in "John loves Mary"), then it will also be possible for "Mary" to be a subject noun phrase and "John" to be an object noun phrase (as in "Mary loves John").

Within the connectionist paradigm, the ability to account for regularities is not tested through representing generalizations (as in the classical paradigm), but through learning generalizations. A connectionist architecture accounts for a regularity if it can generalize from a training set to a testing set in accordance with that regularity. Hadley (1994a) uses this criteria to formalize the concept of systematicity. He requires a network to generalize from a set of training sentences to a set of novel testing sentences. He defines three degrees of systematicity, depending on the novelty of the testing sentences. The one which most closely matches Fodor and Pylyshyn's concept is strong systematicity, where the novel testing sentences include words (i.e. simple constituents) in syntactic positions where they did not occur in the training sentences. Going back to our previous example, this means that a network that was trained on sentences in which "Mary" was only in object position would have to handle sentences in which "Mary" appeared in subject position. Unfortunately, as Christiansen and Chater (1994) point out, this definition still has some imprecision in that no definition of "syntactic position" is given. Christiansen and Chater (1994) provide a linguistically motivated definition of syntactic position, but this is not adequate. The network may be using a rather different system of syntactic positions than the one an external observer would find natural.

For the test to guarantee that the network is truly able to generalize to novel pairings of words with syntactic positions, we need to restrict the network's task so that the syntactic position of the word in the training set must be treated as different from the syntactic position of the word in the testing set. Hadley (1994a; 1994b) discusses a fourth type of systematicity which does this. He defines semantic systematicity to be strong systematicity plus the requirement that the system assign appropriate meanings to all words occurring in the novel test sentences (Hadley, 1994b). The task of assigning meaning forces the system to make distinctions between syntactic positions. For example, a network trying to assign meaning to "Mary loves John" must distinguish between the syntactic positions of "Mary" and "John", since neither word can in general be excluded from either the lover or loved roles. This task illustrates how a network can be forced to represent the distinctions between syntactic positions that the experiment presumes. With this restriction on the task, we can be sure that a network that exhibits strong systematicity represents information about syntactic positions and can combine this information with its information about words in novel ways.

Hadley (1994) explicitly excludes from his definition of systematicity a property which Fodor and Pylyshyn (1988) emphasize, namely that the regularities Hadley discusses must be nomic necessities. In other words, these regularities must be inherent to the nature of the connectionist network, not just wired in. Wiring in the regularities constitutes a mere implementation of them, and thus neither explains them nor furthers our understanding of them. This paper argues that connectionist networks which use temporal synchrony variable binding inherently generalize information about words from one syntactic position to another, and thus that strong systematicity is a nomic necessity given the use of temporal synchrony variable binding, as required by Fodor and Pylyshyn (1988). Because we are concerned with demonstrating an inherent property, this paper provides an in-principle argument. Experimental results from an implemented system are neither necessary nor sufficient to demonstrate that a property is inherent to the system.[3]

Learned Parameters versus Variable Bindings

The above definition of systematicity requires that a network make use of information about words, syntactic positions, and the pairings of words with syntactic positions. The pairings are manifested in the sentences that are input to the network, and the task forces these pairings to be manifested in the output of the network. Thus information

[3]This approach is at odds with standard practise in connectionism, but this divergence is to be expected given that the challenge posed by Fodor and Pylyshyn (1988) is a philosophical one, and not directly empirical. For those who aren't satisfied with an in-principle argument, see the discussion below of (Hadley and Hayward, 1995). Their experimental results can be interpreted as evidence for the generalization ability of temporal synchrony variable binding networks.

about these pairings must be communicated from the input to the output, and consequently this information must be represented in some way by the pattern of activation in the network. This is an instance of the variable binding problem. The network must represent the binding between the information about a word and the information about the syntactic position of that word. In classical approaches these bindings are represented using variables. For example, the binding between "Mary" and the subject syntactic position in "Mary loves John" could be represented (simplistically) as $\mathit{Mary}(x) \wedge \mathit{subject}(x)$.

The requirements systematicity places on the representation of variable bindings are rather different from the requirements it places on the representation of words and syntactic positions. In order for a network to generalize from processing one set of sentences to processing another set, the parameters that are determined using the training set (i.e. the link weights) have to represent information that is also true of the testing set. For strong systematicity, both information learned about words and information learned about syntactic positions will be true of the testing set. In contrast, because words and syntactic positions are paired differently in the training and testing sets, any information learned about variable bindings will not be true of the testing set. Since the task requires information about words and syntactic positions to be learned correctly, the parameters that represent this information must not be dependent on variable bindings. In other words, strong systematicity requires that the learned parameters (weights) of a network should be independent of the variable bindings.[4] Classical approaches use quantifiers to express this independence. For example, the learned fact that "Mary" has the interpretation MARY (following a standard notation) could be represented with the rule $\forall y, \mathit{Mary}(y) \Rightarrow \mathit{MARY}(y)$. The truth of this rule is a learned parameter, and it is independent of the variable that the rule is applied to. Applying this rule to x in $\mathit{Mary}(x) \wedge \mathit{subject}(x)$ we get $\mathit{Mary}(x) \wedge \mathit{subject}(x) \wedge \mathit{MARY}(x)$. This state specifies that the subject has the interpretation MARY, despite the fact that the rule is independent of the pairing of "Mary" with the subject.

It should be clear at this point that the requirements of strong systematicity are not specific to the task of natural language interpretation. Information about words and syntactic positions could be replaced with a wide variety of different kinds of information, and we could still find tasks that require these different kinds of information to be independent from their variable bindings. Any such task will display the pattern of regularities that are embodied in the concept of systematicity. For example, grasping a piece of candy that is in ones visual field requires the information that an object is candy, the information about the location of that object, and the binding between these two types of information. Changing the location of the candy does not change its sweetness, and changing the color of the candy does not change the trajectory for grasping. As Fodor and Pylyshyn (1988) argued, such regularities are pervasive in higher level cognitive activities, and thus a proposed cognitive architecture must account for them. In the rest of this paper we will describe a connectionist architecture in which learned parameters are inherently independent of variable bindings, thereby accounting for these regularities.

Representing Variable Bindings

Several researchers have proposed ways in which connectionist networks can represent variable bindings, but most of these methods do not make learned parameters independent of variable bindings, as required by systematicity. The learned parameters of a connectionist network are represented in its link weights. The effect of a link weight on a computation is dependent on which units the link connects and the units' activation levels, but it is not dependent on the time at which the computation takes place.[5] Thus if the time dimension is used to represent variable bindings, then learned parameters will inherently be independent of variable bindings. This is the approach taken in temporal synchrony variable binding. In contrast, if either the space dimension (different units) or activation levels are used to represent variable bindings, then learned parameters will not inherently be independent of variable bindings. It is possible to hardwire the network in such a way as to enforce this independence, but there is no apparent motivation for such hardwired structure other than implementing systematicity, and thus it is not inherent to the network. Perhaps we have missed a method that would address these criticisms, but the existing alternative proposals for connectionist variable binding do not indicate what it would be. Thus currently only temporal synchrony variable binding implies that learned parameters are inherently independent of variable bindings, as required by systematicity.

Tensor product variable binding (Smolensky, 1990) and relative-position encoding (Barnden and Srinivas, 1991) use the space dimension to represent variable bindings. Using such a representation, variable bindings are represented by specifying the units where the computation should take place. The problem with this method is that without additional mechanisms the weights of the links for one set of units will be different from the weights of the links for another set. Thus the learned parameters used in a computation will be different depending on the binding that is manifested in the input, and the network will not exhibit systematicity. If all links are trainable, then the weights for two different sets of units will only be the same if either there is an additional mechanism for enforcing weight equality, or they are trained on equivalent data. Any mechanism for enforcing weight equality across different sets of units is inherently nonlocal, and thus violates one of

[4] This requirement does not preclude the use of information that is dependent on the pairings of words with syntactic positions. Such information is simply irrelevant to the issue of systematicity.

[5] Of course a link weight will have a different effect if it changes from one time to another. This is not relevant here because learning is taking place at a much larger time scale than individual computation steps. Therefore if any change did occur it would be negligible.

the basic tenets of connectionism. Thus, while it may be an effective engineering solution, weight sharing does not constitute a connectionist method for capturing systematicity. Given that the training set used in the above test for systematicity is by design biased with respect to the bindings between words and syntactic positions, the data used to train links for different sets of units will not be equivalent without additional, as yet unproposed, mechanisms.

Hadley and Hayward (1995) propose a more plausible way of using tensor product variable binding, but as the system is described they are still hardwiring in the independence between learned parameters and variable bindings. The component of their network which uses tensor product variable binding has no trainable links. This allows the different links in the variable binding component to stay equal throughout training, thus making the effects of the learned parameters independent of which of the variable binding units is used. The resulting network generalizes extremely well on the small grammar they use in their experiments. However, this proposal has at least two problems; no independent motivation is given for this hardwired component, and the hardwired component grows with the size of the systems vocabulary (linearly for units and quadratically for links), which for a real system would be quite large. They justify this use of hardwired structure by saying that their units and links should be interpreted as high-level abstractions which aren't necessarily manifested as collections of neurons and synapses. While all connectionist networks are abstract models, without some plausible connection to the biological substrate it is difficult to see how the network could be anything more than a mere implementation of a classical statistical system. Interestingly, this criticism can be addressed by assuming that Hadley and Hayward's work is not a competing proposal with the one advocated here, but is actually a complementary one. As pointed out by Tesar and Smolensky (1994), temporal synchrony variable binding can be interpreted as an implementation of tensor product variable binding.[6] In addition, there is some evidence that temporal synchrony is used in the brain to do variable binding (see (Shastri and Ajjanagadde, 1993) pages 439–441 for a discussion). Thus if we assume that the variable binding component of Hadley and Hayward's network is an abstract representation of a temporal synchrony variable binding mechanism, then the "mere implementation" criticism is addressed. More arguments against this criticism will be given below. In the other direction, the experiments run by Hadley and Hayward (1995) demonstrate that when systematicity is embodied in a network (as in temporal synchrony variable binding networks), simple connectionist learning techniques can be very effective for grammar induction.

Signatures (Lange and Dyer, 1989), CONSYDERR (Sun, 1992), and pattern-similarity association (Barnden and Srinivas, 1991) use the activation level dimension to represent variable bindings. This complicates the nature of computation in the network, since activation level is also being used to represent what features a variable's entity has. It is conceivable that these two types of information could be folded into individual activation levels, but it isn't at all clear how this could result in different variable bindings being treated the same but the presence or absence of features being treated differently. All the above investigations use the alternative approach, in which these two types of information are represented in the activation levels of two different sets of units. In this approach, coordinating computation between the two sets of units requires representing the bindings between the variables and their entity's features. These bindings do not have to be dynamically instantiated, so fixed spatial relationships can be used. However, systematicity still requires learned parameters to be independent of these bindings. Because these bindings are represented using space, they pose the same problems as using space to represent variable bindings.

Temporal synchrony variable binding (Shastri and Ajjanagadde, 1993) is currently the only proposal for how to use the time dimension to represent variable bindings. In this model, a variable binding is represented by specifying the times during which the computations involving the variable should take place. If two units are representing information about the same variable, then they output activation at the same time (i.e. synchronously). For our task, if a given word is in a given syntactic position, then the units that represent the word are outputting activation at the same time as the units that represent the syntactic position. Computations are performed when this activation spreads through the network's links. No matter at what time this computation occurs, the same link weights will be used. Thus no matter what variable binding a word or syntactic position participates in, the same learned parameters will apply to it. In other words, learned parameters are independent of variable bindings. Thus the information that the network has learned about a word or syntactic position in one set of word-position pairings will automatically be applied to the same word or syntactic position in different pairings. Therefore the use of temporal synchrony variable binding inherently results in a network which generalizes information about words from one syntactic position to another, and it inherently exhibits strong systematicity.

Temporal Synchrony Variable Binding

Temporal synchrony variable binding is a technique that can be applied to virtually any style of connectionist model. For higher level cognitive activities such as language interpretation recurrent networks are of particular interest. Recurrent networks accept a sequence of inputs over time, and perform a sequence of computations. As with variable binding, the use of time to represent the input sequence allows the learned parameters of a recurrent network to be independent of absolute position in the input sequence. Thereby such networks can generalize what they

[6] Tesar and Smolensky (1994) argue that the use of time rather than space to represent variable bindings is purely an implementation issue. This is reasonable when, as they do, one only looks at static representations. However, as argued here, when the issue of learning is taken into consideration the use of time rather than space becomes quite important, even at the architecture level.

learn about one position in the input sequence to other positions. This property is imperative for language interpretation, where the absolute position of a word in a sentence carries virtually no information. Thus we need a network that can use the time dimension to represent both input sequence and variable bindings.

All that is needed for a network to represent both input sequence and variable bindings in the time dimension is units that pulse periodically. The periods of the resulting pattern of activation correspond to steps in the computation, and the phases correspond to variable bindings.[7] In effect this method simply time-multiplexes a recurrent network across variables. Such a periodic pattern of activation is illustrated in figure 1, where there are two variables and three computation steps. In the initial computation step the pattern of activation represents the information $\mathit{Mary}(x) \wedge \mathit{John}(y) \wedge \mathit{subject}(x) \wedge \mathit{object}(y) \wedge \mathit{active}$. Then some of the links of the network propagate activation from the "Mary" unit to the "MARY" unit, resulting in the pattern shown in the second and subsequent period. These links implement the rule $\forall z, \mathit{Mary}(z) \Rightarrow \mathit{MARY}(z)$. Information that is not predicated of a variable is represented with units that do not pulse, and thus stay active across phases. Such a unit is shown in figure 1 labeled "active" (for active voice, as opposed to passive voice). These units can be used to represent global context, and to coordinate computation across variables. In effect they are a subpart of the recurrent network that is not time-multiplexed. Any aspect of a task for which systematicity is not applicable can be handled within this subpart.[8]

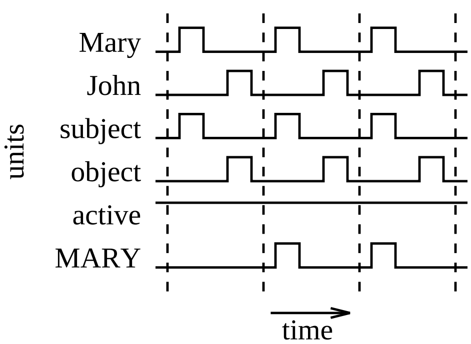

Figure 1: An example of how temporal synchrony can be used to represent variable bindings. Each solid line shows the output of a unit over time. The dashed lines divide this temporal pattern into periods.

The above discussion of temporal synchrony variable binding demonstrates that this technique can be interpreted as an implementation of a classical computational architecture. However, it is not a *mere* implementation. The inherent nature of systematicity in temporal synchrony variable binding networks and the biological evidence for this implementation method means that the pervasiveness of systematicity in cognitive activities is explained by this choice of implementation. This is in contrast to the situation in classical approaches, where compositionality is used to capture systematicity. Compositionality has no independent motivations (other than mathematical simplicity), and thus should be considered a description of the phenomena of systematicity, and not an explanation of it. The biological evidence for temporal synchrony variable binding has also been used to explain other cognitive phenomena that were previously described in classical terms, such as Miller's (1956) bound on short term memory of seven plus or minus two things (Shastri and Ajjanagadde, 1993). Furthermore, the less abstract level of description provided by temporal synchrony variable binding has provided insights into cognitive phenomena that have not been achieved at the classical level. For example, some constraints on long distance dependencies in natural language (wh- movement) can be explained by the inability of simple temporal synchrony variable binding networks to generalize over pairs of constituents (Henderson, 1994b). By accounting for these particular phenomena at the implementation level, the classical competence theory of long distance dependencies is greatly simplified. Because temporal synchrony variable binding helps bridge the gap between low level biological evidence and high level cognitive phenomena, more such insights are likely in the future. Even now it is abundantly clear that temporal synchrony variable binding is not *merely* an implementation of a classical computational architecture.

The parallel between the use of time in recurrent networks and the use of time in temporal synchrony variable binding means that methods for training recurrent networks can be generalized to temporal synchrony variable binding networks. Thus just as a recurrent network can learn that "the" is usually followed by a noun regardless of where it occurs in the sentence, a temporal synchrony variable binding network can learn that "Mary" is usually a noun phrase regardless of what syntactic position it has in the sentence. Some complications arise when it is necessary to learn when to introduce new entities, but these are orthoganal to the issue of systematicity. Methods for effective learning in temporal synchrony variable binding networks is an area of active research by the author and Shastri.

Conclusion

Temporal synchrony variable binding (Shastri and Ajjanagadde, 1993) is a connectionist method for representing multiple entities, each with multiple features. It uses the time dimension to represent the binding between the different features of a given entity (i.e. variable bindings). Because the effect of a link weight on the network's computation is independent of the time at which the computation occurs, the learned parameters of the network are inher-

[7] While the periodic firing of groups of neurons at the same frequency is overly simplistic, it appears to be an appropriate model at this level of abstraction. The only necessary properties here are that groups of neurons can achieve some form of synchronous firing and maintain that synchrony over some period of time, which they can do (Gray et al., 1991).

[8] For a more thorough presentation of temporal synchrony variable binding and how it can be used to implement a syntactic parser, see either (Henderson, 1994a) or (Henderson, 1994b).

ently independent of its variable bindings. In other words, learned parameters inherently generalize across entities. In particular, information learned about words and information learned about syntactic positions will both generalize to new pairings of words with syntactic positions. This is the criteria Hadley (1994a) develops in his formalization of Fodor and Pylyshyn's (1988) concept of systematicity. Thus we have succeeded in showing that temporal synchrony variable binding is a connectionist architecture that accounts for systematicity. In the process, we have clarified the notion of systematicity and shown its implications for variable binding techniques.

While this demonstration has been couched in the terms of natural language interpretation, it is clear that the ability to generalize across entities is applicable to a broad range of higher level cognitive activities. Arguably, it is precisely the lack of this generalization ability that has prevented standard connectionist networks from matching the abilities of symbolic systems in these tasks. The work presented in this paper indicates how the impressive learning abilities of connectionist networks for pattern matching tasks can be extended to the more complex domains typical of higher level cognition.

References

Barnden, J. and Srinivas, K. (1991). Encoding techniques for complex information structures in connectionist systems. *Connection Science*, 3(3):269–315.

Christiansen, M.H. and Chater, N. (1994). Generalisation and connectionist language learning. *Mind and Language*, 9(3):273–287.

Fodor, J.A. and McLaughlin, B. (1990). Connectionism and the problem of systematicity: Why smolensky's solution doesn't work. *Cognition*, 35:183–204.

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

Gray, C.M., Engel, A.K., Koenig, P., and Singer, W. (1991). Properties of synchronous oscillatory neuronal interactions in cat striate cortex. In Schuster, H.G. and Singer, W., editors, *Nonlinear Dynamics and Neural Networks*. Weinheim: VCH Publishers.

Hadley, R.F. (1994a). Systematicity in connectionist language learning. *Mind and Language*, 9(3):247–272.

Hadley, R.F. (1994b). Systematicity revisited: Reply to Christiansen and Chater and Niklasson and van Gelder. *Mind and Language*, 9(4):431–444.

Hadley, R.F. and Hayward, M.B. (1995). Strong semantic systematicity from unsupervised connectionist learning. In *Proceedings of the Seventeenth Conference of the Cognitive Science Society*, pages 358–363, Pittsburgh, PA.

Henderson, J. (1994a). Connectionist syntactic parsing using temporal variable binding. *Journal of Psycholinguistic Research*, 23(5):353–379.

Henderson, J. (1994b). *Description Based Parsing in a Connectionist Network*. PhD thesis, University of Pennsylvania, Philadelphia, PA. Technical Report MS-CIS-94-46.

Lange, T. and Dyer, M. (1989). High-level inferencing in a connectionist network. *Connection Science*, 1(2):181–217.

Miller, G.A. (1956). The magical number seven plus or minus two. *Psychological Review*, 63:81–96.

Niklasson, L.F. and van Gelder, T. (1994). On being systematically connectionist. *Mind and Language*, 9(3):288–302.

Shastri, L. and Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16:417–451.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46(1-2):159–216.

Sun, R. (1992). On variable binding in connectionist networks. *Connection Science*, 4(2):93–124.

Tesar, B.B. and Smolensky, P. (1995). Synchronous firing variable binding is a tensor product representation with temporal role vectors. In *Proceedings of the Sixteenth Conference of the Cognitive Science Society*, pages 870–875, Atlanta, Georgia.

von der Malsburg, C. and Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54:29–40.