Active Symbols and Internal Models: Towards a Cognitive Connectionism

Stephen Kaplan[1], Mark Weaver[1], and Robert French[2]

[1]Department of Electrical Engineering and Computer Science, and Department of Psychology, The University of Michigan, Ann Arbor, MI 48109, USA
[2]Center for Research on Concepts and Cognition, 510 N Fess, Indiana University, Bloomington, IN 47401, USA

Abstract. In the first section of the article, we examine some recent criticisms of the connectionist enterprise: first, that connectionist models are fundamentally behaviorist in nature (and, therefore, non-cognitive), and second that connectionist models are fundamentally associationist in nature (and, therefore, cognitively weak). We argue that, for a limited class of connectionist models (feed-forward, pattern-associator models), the first criticism is unavoidable. With respect to the second criticism, we propose that connectionist models *are* fundamentally associationist but that this is appropriate for building models of human cognition. However, we do accept the point that there are cognitive capacities for which any purely associative model cannot provide a satisfactory account. The implication that we draw from is this is not that associationist models and mechanisms should be scrapped, but rather that they should be enhanced.

In the next section of the article, we identify a set of connectionist approaches which are characterized by "active symbols" – recurrent circuits which are the basis of knowledge representation. We claim that such approaches avoid criticisms of behaviorism and are, in principle, capable of supporting full cognition. In the final section of the article, we speculate at some length about what we believe would be the characteristics of a fully realized active symbol system. This includes both potential problems and possible solutions (for example, mechanisms needed to control activity in a complex recurrent network) as well as the promise of such systems (in particular, the emergence of knowledge structures which would constitute genuine internal models).

Keywords: Artificial intelligence; Connectionism; Human cognition; Symbol systems; Behaviorism; Learning; Knowledge structures

Introduction

There can be little doubt that the connectionist revolution is under way. In the last five years, the class of techniques that fall under the collective heading of

"connectionism" has become the single most credible competitor to three decades of research in artificial intelligence. Originally a response to the failures of the "symbolic" paradigm to live up to the expectations of the sixties and seventies, connectionist networks have shown promise in areas such as object recognition, pattern completetion, speech synthesis, and verb conjugation. However, these achievements do not, in and of themselves, constitute revolutionary advances in artificial intelligence. Their real importance lies not so much in the results, but rather in the means used to achieve them.

Connectionists, for the most part, emphasize "neurologically plausible" approaches to cognition and, as a result, their models use networks of interconnected nodes and simple spreading-activation algorithms rather than the traditional arsenal of artificial-intelligence techniques such as formal logic, heuristic search, explicit rules, and so forth. Another crucial aspect of the connectionist approach is that connectionist networks are typically trained, rather than explicitly programmed, to perform their tasks. These radical departures from the traditional approach justify the expression "connectionist revolution."

But the revolution is by no means won. Given that the premises of connectionism challenge the central dogmas of almost thirty years of work in artificial intelligence, it is hardly surprising that critics abound. We will devote considerable attention in this article to reviewing and analyzing the various criticisms of connectionism. Essential to our comments is the recognition that the connectionist endeavour is not homogeneous. While it is true that all connectionist models share a node-and-link structure, throughout which various algorithms govern the spread of activation and the weights on the links, the similarities essentially end there. For the moment, it would be difficult – probably impossible – for the connectionist community to agree upon some overarching set of "connectionist" principles.

It is our opinion that the tendency to treat connectionism as if it were monolithic has been a source of confusion and error. We feel that it is impossible to make sense of arguments, either for or against connectionism, without a clear understanding of the characteristics of the models in question. With this in mind, we will focus on two major criticisms, namely:

connectionism is merely a naive, computerized revival of behaviourism and

connectionist models are fundamentally associationist in nature, and this severely limits their cognitive potential.

We will see that, with respect to what is currently the most popular class of models, the feed-forward pattern associators (e.g. back-propagation networks), the behaviourism charge is largely justified. We will describe in considerable detail the architecture of an alternate class of connectionist models, based on the recurrent circuit, for which the behaviourism criticism does not apply. We will not contest that connectionist models are associationist but, we will argue that this has improperly been considered a disadvantage only because the power and ubiquity of association in cognition has been underestimated.

This discussion that follows is divided into three sections. First, we consider certain criticisms of connectionism. We will examine, in particular, the charge

that connectionism is naive behaviorism. This will be followed by a detailed examination of the recurrent circuit, the building block for a class of connectionist models that avoid the criticisms presented in the first section. Finally, we will discuss how recurrent-circuit architectures are capable of producing associative structure. We will attempt to show just how powerful systems with associative structure are.

Criticisms of Connectionism

"Connectionism is actually little more than a warmed over version of . . . ?" Choose a discredited, obsolete psychological theory to finish the sentence and a critical analysis is off and running. Two popular choices for this purpose seem to be "behaviourism" and "associationism". Are these characterizations accurate? Before we can address this issue, we need to point out that, while for the modern reader, "behaviorism" and "asociationism" might seem, superficially, to be synonymous, in fact the two terms have very different implications.

Both the 19th-century associationists and the 20th-century behaviorists are distinguished by an emphasis on learning and by a belief in the importance of contiguity and repetition in that context. The critical difference between the two, however, lies in the respective building blocks of association. The only associations that behaviorists considered were those between observable entities, that is between stimuli and responses. The associationists, on the other hand, were concerned with the association of "ideas". In spite of the fact that "idea" was a rather nebulous construct by modern standards, it clearly referred to an internal, mental entity. So, while behaviorism denied the existence (or at least the usefulness) of cognitive processes, associationism was a cognitive theory.

What perspective does this provide to the "connectionism equals behaviorism" charge? First, it is clear that modern connectionists are not behaviorists at the conceptual level. That is, modern connectionists are proposing an alternate way of modelling cognition; they certainly do not deny the existence of cognition or the validity of studying it. However, we cannot dismiss the charge of behaviorism quite so easily. The possibility remains that the connectionist models themselves have unavoidable behaviorist implications which their proponents did not intend. We will argue that this, in fact, is the case, but only for a limited (although popular) class of models. We will discuss a class of connectionist models which are not at all vulnerable to the behaviorism characterization.

The argument surrounding "connectionism equals associationism" is more complex. Association is a fundamental property of connectionist models, but we will argue that this is not the disadvantage that it is sometimes made out to be.

Connectionism Equals Behaviorism

It is our contention that the behaviorism characterization applies only to a limited class of models which we will refer to as feed-forward pattern associator (FFPA)

models. FFPA models are those which are trained to associate arbitrary sets of state vectors, and we include in this class both those models which lack layers of "hidden" units and employ the standard perceptron learning rule (or delta rule) as well as models which do possess hidden units and which employ the more sophisticated "back-propagation" learning rule (Fig. 1). The focus on FFPA models is appropriate for several reasons. These models are the subject of ongoing and widespread study within the connectionist community.

FFPA models have been used to produce some of the most striking and best-known results of modern connectionism is such areas as speech synthesis (Sejnowski and Rosenberg, 1986) and past-tense verb learning (Rumelhart and McClelland, 1987). As such, they have been the basis of some of the most revolutionary claims made by the connectionists. Finally, there are even widespread efforts to find commercial applications for such models as, essentially, "trainable" expert-systems. Thus, whatever their limitations, understanding the properties of FFPA models is important.

The case for "connectionism equals behaviorism" can be found explicitly in Lachter and Bever (1988) and implicitly in Prince and Pinker (1988). Both papers include detailed examinations of Rumelhart's and McClelland's model of past-tense verb learning. Considerable effort is devoted to explaining why this model only gives the appearance of learning the past tense of English verbs in a manner that mimics the way children do. Their conclusion is that the success of Rumelhart and McClelland's model is not genuine, but is rather due to certain *ad-hoc* aspects of presenting the input to the model. What is important for our discussion is that they go on to argue that FFPA models are inherently incapable of the past-tense production task because of the fact past-tense production must be implemented as a direct mapping from stem to past-tense form. This direct mapping means,
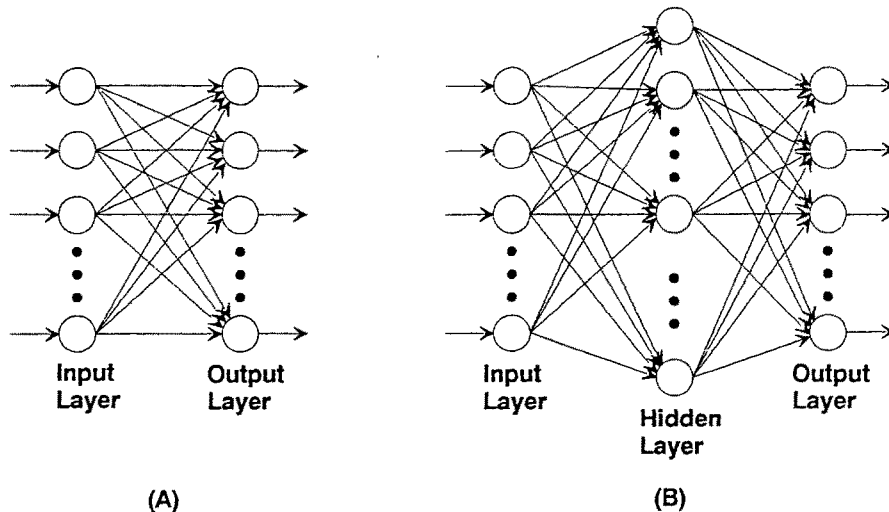

Fig. 1. Feed-forward, pattern-associator models. (A) a simple, 2-layer pattern associator. (B) a multilayer, back-propagation network.

according to Lachter and Bever, that FFPA models are actually complex, parallel versions of a traditional S–R model.

What are the consequences of the direct mapping, of an S–R approach? Prince and Pinker enumerate the following deficits: the lack of a lexical item, the lack of modular operation, an inability to represent and use context. Consider the situation where two distinct verbs share the same stem, for example "The neighbor rang the doorbell" and "The army ringed the city". The past-tense learning model cannot handle this because there is no unique mapping from stem ("ring") to past-tense form ("rang"/"ringed"). In order to handle such verb pairs correctly, a system must, at minimum, have an internal representation (a lexical item) for each of the alternatives. These representations would then mediate the past-tense mapping (creating modularity). Furthermore, to select between the alternatives, a system would have to employ information which is of a different nature (semantic as opposed to phonetic) and which is temporarily separated. That is, a system must have distinct internal representations for each verb stem, and it must be able to select between them based on an internally maintained context.

Are FFPA Models Behaviorist?

What defense can one make to the charge that FFPA models are inherently behaviorist? On the surface, there are some rather obvious differences between FFPA connectionism and the behaviorist theories of 40 years ago. In behaviorist learning theories, the S's and R's were atomic, and a separate, independent link was required for each S–R association. Also, following Lashley's famous series of experiments (Lashley, 1950), which failed to find any physical link between S's and R's, the neurophysiological mechanics of learning were not considered a suitable subject for research. In contrast, the S's and R's of back-propagation networks are feature patterns, and the associations between all S's and R's are superimposed on the same network; this yields simple but powerful generalization capabilities which eluded the behaviorists. Notably, too, the mechanisms by which the S's and R's are linked, far from being off-limits, are the subject of intense scrutiny (see for example, Rosenberg, 1987).

However, despite these differences, there is a conceptual commonality between FFPA models and behaviorist psychology which overshadows these other considerations: both systems are completely environmentally driven. That is, the responses of the behaviorist organism and of the feed-forward network are fully determined by the stimulus. In neither behaviorist nor FFPA models is there an allowance for, or a possibility of, internal cognitive processes.

Of course, in examining the past-tense learning model we have been discussing the least powerful type of FFPA. Specifically, the past-tense learning model lacks the layer of "hidden" units found in a back-propagation net (Fig. 1). How does the addition of hidden units affect the behaviorism argument? Rumelhart and McClelland argue that hidden units are crucial:

In our models, we are explicitly concerned with the problem of internal representation and mental processing, whereas the radical behaviorists explicitly denied the scientific utility and even the validity

of the consideration of these constructs. The training of hidden units is . . . the construction of internal representations.

But this argument will not stand up to scrutiny. The connections between the input and output layers in a back-propagation network are no longer direct, but are mediated by one or more layers of "hidden" units. This addition of hidden layers indeed adds computational power; such networks can, for instance, learn to solve the infamous "exclusive-or" problem which is impossible in two-layer networks (Minsky and Papert, 1969). Unfortunately, all of this is irrelevant to the issue of behaviorism. Although the connection between the input and output layers is no longer direct in terms of network architecture, it remains direct in terms of time and causality. When a stimulus is presented to the input layer, a response follows immediately which is fully determined by that stimulus. This property, the complete determination of each response by the characteristics of the immediate stimulus, is a more accurate description of the behaviorist position, and it is true of FFPA networks regardless of whether or not they contain hidden layers.

Connectionism Equals Associationism

Fodor and Pylyshyn (1988) present a detailed argument that connectionism equals associationism and that, as a result, connectionism has serious limitations as a cognitive theory. Fodor and Pylyshyn do concede that "both connectionist and classical architectures postulate representational mental states". Thus they are not claiming that connectionist models lack internal symbols and internal processing. Instead, they claim that although connectionist models may have symbols, the only mechanism of interaction between symbols is association. Therefore, according to Fodor and Pylyshyn, connectionist systems lack the means of composing symbols into larger structures and they lack operations to manipulate such structures.*

This argument works best in the context of language and logical inference. Fodor and Pylyshyn point out that language abilities are "systematic", that being able to produce or comprehend one sentence ("John loves the girl") should imply the same capacity for another combination of the same elements ("The girl loves John"). Notice that this argument refers to the ability of language users to re-organize linguistic "atoms" (words) into a virtually infinite number of utterances. So the requirement here is for a generative capacity.

Because Fodor and Pylyshyn wish to conclude that all minds are properly characterized as classical symbol systems, they realize that their success also depends on being able to extend the argument beyond language and logic to cognition as a whole. There are serious problems, however, in this effort. Fodor

---

* An unavoidable complication of this discussion is that the term "structure" is used in two different ways. The first usage is structure as the syntactic structure of a sentence, and the second is structure as in the causal or predictive structure of an environment. As we will argue, these kinds of structure require different kinds of cognitive capacities.

and Pylyshyn provide the following as an example of what it would mean for an organism to be non-systematic:

though you could teach the creature to choose the picture with the square larger than the triangle, you couldn't for the life of you teach it to choose the picture with the triangle larger than the square.

We note that there is no generative requirement posed by this problem; there is no need for the organism to re-organize its "cognitive" atoms in real-time. With language, the need is to be able to organize words, spontaneously, into a unique utterance (structure) which may never have existed before. In the animal-learning example, there is a need for systematicity, but it is a systematicity of a different kind. Here, the issue is for the organism to model an arbitrary causal relation in its environment. In like fashion, the challenge for a connectionist system is to assure the availability of connections for that purpose. This is not a trivial problem, but it has been addressed within the connectionist framework (Feldman, 1988).

Fodor and Pylyshyn's failure to distinguish between different kinds of systematicity reflects a similar failure with respect to mental structure. The structure Fodor and Pylyshyn refer to in their language and logic discussion is that of a sentence under construction or a proof in process – there are structures which are internally generated and transformed in real-time. This is in contrast to learned structure which constitutes a model of causal relationships in the environment. We will argue that such structures are extremely powerful and that connectionist approaches are well suited to developing them.

Within the limited domains of language and logic, however, Fodor and Pylyshyn's case for a generative capacity, systematicity, structure-sensitive operation has some force. Consequently, we will offer a glimpse of where a connectionist ought to look for mechanisms which can handle these tasks. This is very preliminary, of course, but it is much less embarrassing a shortcoming than it would be if classical symbol computation models themselves had demonstrated anything resembling a mastery of natural languages.

The Active Symbol

The fundamental building-block of the alternative class of connectionist models that we describe in this article is the recurrent circuit as representation. By recurrent circuit, we mean a "cyclic" network in which activity can persist over time (reverberation). The recurent circuit is an idea whose origins are found in the Hebbian cell-assembly (Hebb, 1949). In this article we will call recurrent circuits which represent concepts "active symbols" after Hofstadter (1979).

Because of the primordial role of the recurrent circuit in the class of models we wish to describe, let us spend a moment reviewing why Hebb first introduced the notion of cell assembly. During the first three decades of behaviorism, not only were many psychologists deeply frustrated because it had become wholly unacceptable to talk about cognition in any terms other than stimulus-response, but, in addition, a number of anomalies with the behavior paradigm had begun to

accumulate (Gibson, 1941; Tolman, 1948). This forced the behaviorists to invoke a number of apparently *ad hoc* constructs to explain this anomalous behavior. Hebb, in a sub-chapter significantly entitled "Rejecting the Assumption of a Complete Sensory Control" (Hebb, 1949), proposed that these construct could be explained only by independent, central processes (see also Milner, 1957). Denying central processes, Hebb argued, was not a viable option; psychologists had to address the issue. He saw a way to do this rigorously by taking into account recent neurophysiological advances, in particular, the discovery of recurrent neural connections and constant brain activity. He therefore proposed the cell-assembly, a recurrent circuit capable of supporting reverberation, as the basis of his connectionism. In this way, behavior would not be uniquely determined by immediate sensory stimulation, but could also be affected by internally persisting activity. Because Hebb's cell-assemblies were proposed explicitly as the basis for central processes, they were a challenge to the behaviorist paradigm. Hebb's work, in fact, can be seen as instrumental in helping break the grip that behaviorists had on psychology at that time, providing a basis for the cognitive revolution that was to follow.

Specifically, the cell-assemblies Hebb proposed were inter-connected clusters of neurons in which activity could persist. Because of this persistence of activity, an active cell-assembly could cause other cell assemblies to become active; activation, in other words, did not have to be exclusively the result of environmental stimulus. Recurrent circuits provide the system with the means – essential, in our view, to any model capable of cognition – of having internal, semi-autonomous, activatable representations of reality (i.e. active symbols) that do not rely uniquely on the sensory interface to become or to stay activated. Recurrent circuits serve as the basis for recognition; they provide a means of coding objects in the environment. Perceptual learning and recognition are not alone sufficient to account for the full range of cognition. A structuring of these active symbols codes objects in the environment is also necessary. The alternate class of connectionist models we describe in this article uses these active symbols as the building blocks of more complex internal models of the environment (see Higher-Level Processes). Even though these models may still not be able to achieve all of cognition, the possibility of long-term associative structure conferred on them by their recurrent-circuit design will, nonetheless, make them very powerful.

Certain connectionists currently exploring recurrent-circuit architectures tend to treat recurrent-circuits as a technical issue. For these researchers, recurrent circuits are a means of achieving greater computing power (Williams and Zipser, 1988). They are not, however, the conceptual basis of their models. On the other hand, a significant number of connectionists have proposed the primacy of the recurrent-circuit approach to connectionism, variously referring to their recurrent circuits cum active symbols as "object representations" (Kaplan and Kaplan, 1982), "recognition codes" (Grossberg, 1987), "classification couples" (Edelman, 1987) and, like Hebb, "cell-assemblies" (Braitenberg, 1978; Palm, 1984).

One important consequence of active symbols is that they provide a means of avoiding the pitfalls of behaviourism. Later in this article we will see how symbol composition can potentially lead to the development of larger structure. How-

ever, like all good things, the recurrent circuit comes at a price. In the following section we will examine three major potential difficulties of recurrent-circuit architectures and the control mechanisms required to solve these problems.

Active Symbols and Control Mechanisms

Three potential problems with a recurrent-circuit design are:

1. What prevents the unchecked spread of activation throughout the net?
2. How does the system select one of a number of mutually exclusive, partially active representations?
3. What prevents a symbol, once it has become active, from remaining active indefinitely?

Fortunately, only two control mechanisms, inhibition and fatigue, are needed to prevent these problems. Let us first consider inhibition. Inhibition, whether it acts locally or over more extended regions, is an activity damping mechanism. When Hebb wrote *The Organization of Behavior* in 1949, the neurological plausibility of inhibition was still hotly debated and, as a consequence, he did not include it in his theory of learning. By the mid-fifties, however, sound neurobiological evidence for inhibition had been found (Eccles, 1957; Milner, 1957). In addition, the results of an early computer simulation of a neural network with a Hebbian learning rule (Rochester *et al.*, 1956) also argued strongly in favor of the necessity of inhibition.

The role of inhibition is crucial in the selection of one representation among two or more partially active ones. Since, for example, the representations of persimmon, tomato, and orange may all become active at the sight of an orange fruit, the organism must have a mechanism allowing it to "select" one of the representations and exclude the others. As soon as one of the representations becomes even slightly more active than the others, it begins to inhibit the activity in the other closely associated representations. This, in turn, accentuates the difference in activity between the inhibitor and its inhibited competitors. This further enhances the inhibitory effect on the less active representations by the most active one. Very soon, one representation completely dominates the others and it is the one that is perceived. This mutually inhibitory effect of similar processes has been documented by Kinsbourne (1982).

The adaptive benefits of such a selection mechanism for representations that are simultaneously partially active is clear. Perfect information and sufficient time to optimally analyze it are, unfortunately, rare commodities for any organism functioning in a real environment. As such, the benefits of rapid selection outweigh the disadvantages of not always picking the optimal representation. The mechanism of inhibition provides just such a means of the rapid selection of similar representations.

We now come to fatigue, the temporal counterpart of inhibition. Inhibition, as we have seen, damps the unchecked spatial spread of activation. Fatigue damps activity over time. The higher the activity in a circuit and the longer the amount of

time that circuit has been active, the greater its fatigue. As a circuit fatigues, its ability to pass activation to its neighbors decreases. Fatigue may be thought of as self-inhibition, the means by which an active representation shuts itself off. Were it not for fatigue, activity in a given circuit might circulate forever. Since associated representations become fatigued in the order in which they became active, fatigue is one means of achieving temporal sequencing of representations. We will return to the issue of sequencing in the section on higher level processes. (For a presentation of a model that makes use of this interplay of inhibition and fatigue, see French and Weaver (1987).)

Symbol Formation and Properties

In systems whose building-blocks are active internal representations, the means of acquiring these representations is a critical issue. Conceptually, the recurrent-circuit connectionists follow Eleanor Rosch's notion of "cutting the world at its joints" (Rosch, 1976). That is, active symbol formation is an unsupervised process of neural response to clusters of correlated features in the environment. This process is generally described as consisting of two levels. At the first level, recurrent circuits corresponding to features are created as units with similar "receptive fields" (and, therefore, correlated activity) have their interconnections strengthened. This process is described explicitly in Braitenberg (1978, 1984) and Edelman (1988). At the next level, the representations of correlated features are linked to form concept representations (or active symbols).

Two primary functional properties of these active symbols warrant description, namely: generalization and the use of activity level as a code. "Generalization" refers to the fact that an active symbol implements a category, responding best to the central tendency of the category and proportionately less well to variants. This means that the active symbol generalizes; it responds to a group of similar stimuli. With generalizations, the concern is with the range of stimuli to which an active symbol will respond. The use of activity level as a code, on the other hand, involves the nature of the response. In general, the level of activity within a symbol codes a degree of matching. Perceptually, this matching is between the stimulus and the category which the active symbol represents. In this sense, activity-level coding can be seen as a means of implementing the 'graded-structure' of categories (Smith and Medin, 1981). Activity-level coding has also been employed by Grossberg (1987) in a model of discrimination-learning, and it has been proposed by Kaplan and Kaplan (1982) and Palm (1982) as a means for testing potential solutions in problem solving.

Higher-level Processes

The active symbol in itself confers considerable power on a connectionist system. This concept makes it possible to put to rest the fears that association-based theory inherently leads back to behaviorism. The centrality and usefulness of active symbols, however, cannot obscure the fact that this level of processing is by

itself insufficent to account for much of intelligent behavior. Some sort of "higher-level" processing is required.

The active symbol plays a central role in this higher-level processing, namely, that it constitutes the basic unit out of which larger cognitive structures, such as patterns coding for temporal and spatial sequence, are created. It is important to note that while the creation of these larger structures is a potential of recurrent-circuit connectionism, at present these possibilities have gone largely unexplored.

The next section concerns these larger structures, with particular emphasis on how they are constituted and how they function. Before beginning this analysis it is perhaps useful to examine what this associative hardware achieves and what it fails to achieve.

As the organism experiences environmental sequences, corresponding internal sequences are created. The active symbols function as elements in these internal sequences. Through associative learning these sequences come to form a network that constitutes a first-order knowledge structure. The achievement at this level is the construction of internal models of the environment.

In the context of a layered system, these models form the input for associative structures that are not models of the environment, but models of models. Such models not only greatly enhance the degree of abstraction possible,, but also make possible the emergence of a powerful reflective capability, namely the capacity of the system to model itself. This cognitive domain has received considerable attention in recent years under the label of "metacognition" (Weinert and Kluwe, 1987).

What remains to be achieved are structure-sensitive operations. Such operations may not be as content free as is often asserted. Further, it is possible that many of these may be spatial in origin. An encouraging start on this challenging problem has been made by physiologists studying locational and parietal processing, and by theories of functioning among these structures.

First-order Knowledge Structures

The sequence is frequently overlooked in discussions of the cognitive process. It does not appear in the index of the vast majority of cognition texts. Yet the capacity of a cognitive system to predict, to plan, and to explore alternative courses of action all depend upon sequence.

Consider problem solving, frequently taken to be paradigmatic of higher-level cognitive processes. It is generally agreed that problem solving is appropriately conceputalized as a search of a problem space. Let us examine what such a space must contain. It presumably includes the important elements of the problem, drawing on both the current situation and the desired end-state or goal. Further, these elements must be organized in some way; some are closer to the goal, others to the current situation, etc. And finally, the elements must be connected in some way so that one can explore possible alternatives, looking ahead to implications of particular courses of action.

A problem space is, in other words, a knowledge structure with the all-important feature of being able to capture the causal links between events

(Macphail, 1987). Such a structure will necessarily be based on experience with environmental sequences. Such sequentially based knowledge structures are central not only to problem solving, but to other essential cognitive processes as well. They are required for foresight (Braitenberg, 1984). Planning, as character-ized by Miller, Galanter and Pribram (1960), operates through the execution of managed sequences. Decision making, where evaluating the likelihood that a given choice will bring one to a given goal is an essential element, also depends upon a sequentially based structure (Kaplan, in press).

The emphasis on causal structure as essential to high-level cognitive activity is, of course, not new. Tolman and Brunswick (1935) long ago pointed to the functional importance of learning the "causal texture of the environment." More recently, Kruglanski and Ajzen (1983) have stressed the centrality of acquiring a causal framework. It is important, however, not to view causality too narrowly. Knowing what follows what is adaptively useful whether the relationship is literally causal or not.

Although the sequence has often been ignored in discussions of cognitive functioning, it has not been completely overlooked. Both Miller (1963) and Macphail (1987) have explicitly linked sequence and causality. Macphail goes so far as to consider the learning of sequence as basic to all intelligent behavior. The sequence as a key aspect of mental functioning is also central to both Shepard (1975) and Kosslyn (1987). It should be noted that knowledge structures based on experienced sequences have long been the concern of theories of associative learning.

Fodor and Pylyshyn's questioning of the significance of associative mechanisms in cognition may be due in part to the tendency to think of association in a behaviorist context. As Voss (1969) has pointed out, although association has long been a central theme in psychology, there has not been agreement as to what it is that is being associated. Using the active symbol as the "element" to be associated makes a substantial difference in how an associative system would function. In particular it counters the common conception of associations acting singly – as indeed they did in most behaviourist theorizing. In a network made up of active symbols, by contrast, the associative connections would be expected to exert their influence through a complex pattern of interaction and convergence.

A simple example of such convergence is represented by the normative reaction to the question 'What do cows drink?' The answer that most commonly comes to mind first is "milk". In this example, due to the French researcher Jacques Pitrat, the answer is so strong because of activation spreading from both "cow" to "milk" and from "drink" to "milk". For those desiring more formal evidence of the power of associative influences in thought, pertinent literature includes both that dealing with the role of spreading activation in language (Meyer and Schvaneveldt, 1971) and the work demonstrating the fundamental role of sequence in the organization of scripts in memory (Barsalou and Sewell, 1985; Abott, Black and Smith, 1985).

In understanding the relationship of sequence to higher-level cognitive proces-sing in the class of connectionist approaches we are considering, it is necessary to distinguish the sequence as an environmental pattern from the sequence as a stored structure. Although environmental patterns that are sequential will lead to

stored structures, non-sequential environmental patterns will also be tend to be stored as sequences. We will argue that temporal sequences provide the means of storing both of these types of patterns.

It is not surprising that most causal sequences should code as temporal sequence. The intimate relation of causality and temporal sequences is not, of course, a new one. David Hume (1737) claimed that it was indeed impossible to discover "any quality which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find that the one does actually, in fact, follow the other.' Somewhat more surprising, perhaps, is that spatial sequences can also be coded in much the same way. Consider, for example, an organism moving through its environment. The limits of that organism's sensory capacity means that only a small portion of the environment will be processed at any one time. Consequently, a particular sequence of (spatial) movements through the environment will be coded as a temporal sequence. One curious, but helpful way to view this situation is to interpret the organism's movement through a stationary environment as the environment moving past the stationary organism. Under those circumstances, the sequences of events (i.e. the moving environment) could be viewed as a causal chain of events and would, as we have already seen, be coded as a temporal sequence.

The same basic connectionist learning rules that lead to the formation of active symbols at a lower level lead to the formation of connections between symbols at this higher level of cognitive organization. As an environmental sequence is experienced repeatedly, a corresponding internal sequence is learned through the connection or association of symbols. Just as experiences with many instances of an object leads to a generic concept of that object, experiences with many particular sequences leads to a connection pattern expressing the characteristic sequential relationships among a set of objects. This connection pattern among symbols constitutes the knowledge structure of the system. It might seem that this knowledge structure would be a collection of sequences. This would be a misleading description; the resulting structure is considerable more interesting (and more powerful) than that. The emergence of this more complex structure, in turn, is a consequence of the pattern-recognizing function of the active symbols. Perhaps an example would be helpful here. Consider the sequences routinely followed by a hypothetical individual whom we can call "John". Let us assume that John experiences the following sequences during a typical week:

1. Home → work → home
2. Home → work → tavern → home
3. Home → work → tavern → auto parts → home
4. Home → auto parts → home
5. Home → auto parts → grocery → home

These different sequences would not be entered into the knowledge structure as separate, unconnected entities since they share elements in common. The result of learning these intersecting sequences would be the cognitive network depicted in Fig. 2. Thus, to return to our example, although he has never experienced it, if John were asked to stop at the grocery on the way home from work, we would
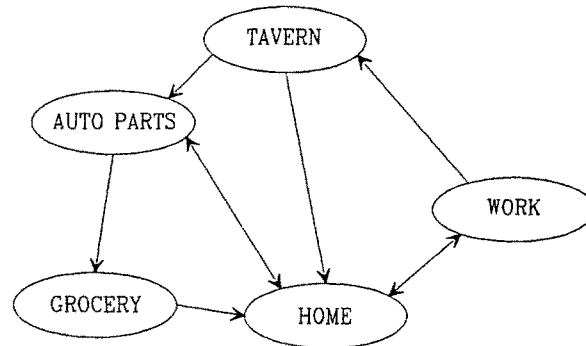
Fig. 2. Superimposing intersecting sequences leads to the beginning of a more complex structure.

assume he would carry out this sequence without difficulty. (Note: This example is discussed in greater detail in Kaplan and Kaplan, 1982, pp. 41–45.) In this way the proposed mechanism is economical not only in the way it stores information, but also in the way it utilizes experience. Thus, through multiply intersecting sequences, there emerges a many-dimensional structure based on experiences of restricted dimensionality. Many of the sequences in this emergent structure will be novel in the sense that they have never been experienced by the organism. They provide an opportunity for discovery if the structure is run off-line, or an apportunity for "responsible innovation" (Kaplan, 1976) if it is functioning to guide the organism's behaviour.

This inclination for intersection among sequences leads to a structure that is essentially an internal map of the experienced patterns in the environment. Such a structure is appropriately considered a cognitive map (Kaplan, 1973; Stea, 1978). The term "map" suggests a certain analogical quality of the knowledge structure, and this suggestion is intentional. The cognitive map tends to correspond to the structure of the experienced environment in the sense that:

1. Sequences in the environment correspond to internal sequences

2. Intersections also correspond

3. What is near in the experienced environment is also near, in terms of the chain of associated symbols, in the corresponding cognitive map

This partially analog property of an active symbol-based associative knowledge structure has a number of interesting consequences. It provides a mechanism for context in the sense that it facilitates retrieval for information related to whatever is currently being perceived. Such facilitation takes place because what is experientially related will also be associatively close and as such will receive facilitation from current activities. Storage of information in this class of systems is thus highly structured. This is largely due to the routing of information at the time of storage, rather than to later reorganization. A quick, albeit oversimplified

look at a consolidation-based learning process will perhaps make clear how coherent structure can arise in an essentially automatic fashion in a connectionist system.

An environmental sequence, when perceived, will give rise to activity in the corresponding active symbols. Symbols at points of sequence intersection will, of course, also become active. The activity is the basis of learned associations in a consolidation-based learning process. Thus new sequences will be "fit into" patterns of old sequences to the extent that they intersect. They will, in other words, be stored in associative proximity to related material, greatly simplifying retrieval at a later time. Since structure is central to many aspects of human cognition (Norman, 1980) and the structure laid down at the time of learning apparently plays a key role in the problem-solving process (Posner, 1973), the automatic generation of coherent structure is an asset of some significance.

Such a mechanism also speaks to the concern raised by Lindsay (1988) who pointed out that within the behaviorist context there is no way to add new knowledge so that it builds upon and can be synthesized with preexisting knowledge. Although he indicated that this remains an unsolved problem for connectionists, the integrating and structure generating tendencies of the active-symbol approach demonstrate how this important concern can be resolved within an associative framework.

The purpose of the cerebral cortex has been described as building internal models of the environment (Barlow, 1985). An internal model would seem to be a highly structured form of information, which is analogous to the structure of the environment in certain key aspects. Such an internal model would permit a variety of essential cognitive processes, presuming that it could be run off-line and that it could be run faster than real time. Such a model would support not only alternative predictions of the future, but assessments of the attractiveness of these different alternatives as well. This is compatible with the view of the mind that suggests that selection of one sequence out of many is a primary activity of the human brain (Calvin, 1987).

It thus appears that the structural and functional requirements for building an internal model are met by the cognitive-map concept. The simplified, partially analog structure described above provides an appropriate structure for embodying the internal model of the world. The inclination to construct such potentially flexible and powerful cognitive maps is an emergent property of this class of connectionist systems (cf. in particular, Braitenberg, 1984 and Kaplan and Kaplan, 1982). Since a number of essential cognitive activities depend upon a landmark-emphasising, organized, connected structure in which sequences are preserved and in which novel sequences can be explored, The cognitive-map concept is an important tool in modelling these higher level processes.

Second-order Knowledge Structures

Although cognitive maps and the internal models they make possible lend considerable cognitive power to an active symbol-based connectionism, such a

scheme is limited in its power of abstraction, and hence in the advantages that abstraction brings. Incorporating this aspect of cognitive functioning is, however, readily achievable by this approach. If one conceptualizes the cognitive system as a layered structure with only a limited portion close to the sensory interface and other portions increasingly distant, then the first layer would be the likely locus of comparatively literal models of the world. Subsequent layers, however, since they would receive their input from prior layers, would build models based on this filtered input. Thus the system would be disposed to build models of models. At some level, such a system is no longer modelling even a highly abstract version of the environment; rather it is modelling its own functioning. In this way it becomes capable of contemplating its own cognitive process; it becomes capable, in other words, of metacognition (Weinert and Kluwe, 1987).

Toward Structure-sensitive Operations

The domain of structure-sensitive operations is a fascinating one. It is also highly seductive. It promises the capability for processing based solely on structure, without any regard to content. The extent to which this is truly characteristic of human cognition is still unclear. Humans are far more facile with logical operations on familiar than on unfamiliar content (Nisbett and Ross, 1980; Johnson-Laird, 1981). (Even sorting playing cards by suite is substantially slower when the colors of the cards are reversed.) Humans do not follow logical rules in their use of negation (Wason and Johnson-Laird, 1972). ("Don't think purple!") Expertise is notorious for its lack of transfer to even relatively similar content domains (Kleinmuntz, 1968; Voss, Tyler, and Yengo, 1983). For further examples of this kind, see Oaksford, Chater, and Stenning (this journal).

Despite this apparent impurity, structure-sensitive operations are nonetheless important to explore and to be able to incorporate into a theoretical framework. Although Fodor and Pylyshyn have difficulty coming up with a non-linguistic example of such an operation, there exists such an operation that has been extensively studied and that has considerable functional significance. This is the operation of mentally rotating objects. Pylyshyn (1980) had at one time asserted that such operations are actually linguistic, but this improbable contention appears to have been thoroughly discredited (Farah, 1988).

Indeed, rather than attempting to linguicize objects in space, it might be far more fruitful to explore the opposite approach, namely that it is the spatial operations that are the more basic. There is in fact evidence that even basic syntactical operations have a spatial basis (Luria, 1973). Further, Lakoff and Johnson (1980) have identified an extensive spatial framework inherent in language.

An encouraging start on this challenging problem has been made by physiologists studying locational and parietal processing (Wise and Desimone, 1988; Lynch et al., 1977) and by theories of functioning involving these structures (Kaplan, 1970; Kosslyn, 1987). The "mind's eye" (Jonides, 1980) presumably operates in a framework of this kind.

Summary and Concluding Remarks

The Continuum of Cognitive Models

In this final section we wish to note that the cognitive models we have been considering form a continuum. We will take as our starting point on this continuum the strict behaviorist models on the first part of the twentieth century.

As we have seen, the behaviorist postion, largely discredited today, was quite clear: Cognition could be analyzed, and ultimately undertstood, by considering stimulus-response alone. The importance, and even the existence in many cases, of any cognitive processes was explicitly denied. Among a number of other difficulties, these models had a particularly hard time providing any reasonable account of stimulus gneralization.

The next class of models on this continuum, the feed-forward pattern associators, have made significant progress with the problem of stimulus generalization. For this reason alone they represent a major advance over the strict behaviorist position. In addition, the proponents of these models, unlike their behaviorist forebears, are interested in cognitive processes. Furthermore, the resounding success of some of the FFPA programs has, after a lapse of a decade and a half, once again focused a great deal of attention on emergent models of cognition. These models, for all their successes, have nonetheless come under fire as being naively behavioristic because they can only be driven by the environment. If input from the environment stops, they stop.

Next come the so-called recurrent-circuit models (or "active symbol-based models") that build internal representations capable of maintaining a certain level of activity independent of the external environment. The recurrent-circuit architecture of these models is what allows active symbols to be produced. The existence of active symbols as building blocks also allows the emergence of associative structure which, in turn, provides a means of coding both the causal structure of the environment and the hierarchical structure of concepts.

The final class of models are those capable of developing structure sensitivity, allowing them to do, in particular, formal logic and certain linguistic operations. Although there is reason to believe that such capabilities have an origin in spatial cognition (Lakoff and Johnson, 1980), at present no detailed theories exist of how this might be done. It is, however, important to add that we feel far too much emphasis has traditionally been placed on the importance of these formal logic capabilities. In human cognition, these operations must be viewed primarily as an enhancement to a primarily associative system and not, as traditional AI has always done, the other way around. Evidence has shown that even in many cases where formal logic would be appropriate, people have a difficult time using it (Nisbett and Ross, 1980; Kahneman and Tversky, 1973; Margolis, 1987).

Logic, Difficulty, and Adaptation

The brain can be viewed as a builder of approximate models that work in a world of limited predictability. For such a system the capacity to generalize experience

and to make good guesses is central. Functionalists of whatever persuasion see the complexity of the environment and the need to act on limited information as pivotal themes in understanding natural systems. Thus the very conceptions of what cognition is and what it is for are as different as are the proposed mechanisms. The functional and the computational perspectives were not always so far apart. Kenneth Craik many years ago took a functionalist perspective in anticipating the power of computer models when he pointed out:

If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, an in every way react in a much fuller, safer, and more competent manner to the emergencies which face it. (Craik, 1943, p. 61)

The contrast between this conception of cognition and one that emphasizes language and logic is striking. While the basic inferential statement "if a then b" has a distinctly predictive flavor, prediction of extended alternative strings of possible behavior and consequent patterns in an environment where there are many unknowns and many uncertainties hardly lends itself to logic. It does, however, fit very well with the sorts of patterns one might expect of a stochastic, sequence-storing, associatively based system.

The conclusion that logic may not after all be that central to understanding human cognition perhaps should not have come as a surprise. As Lachman and Lachman (1979) have pointed out, a system that was invented as a corrective for human thought constitutes an improbable candidate for being the basis of thought.

In any proposed cognitive system some functions are easy and others are difficult; as Smolensky (1988) has put it, with any given structure there are some things "one gets for free". By ordering cognitive processes along a dimension from easy to difficult, one could construct a hypothetical difficulty gradient. Such a gradient could serve as an important index of the adequacy of the proposed system as a model of human functioning. In an active symbol system, for example, associations are achieved easily, in fact so easily as to provide default solutions when other operations fail or are not available. Logic, by contrast, is difficult. It is our contention that this is a far better description of the way that humans think than is the opposite-leaning gradient favored by Fodor and Pylyshyn. Thus, the opportunity remains for connectionist approaches to continue to develop an account of cognition that is neither behaviorist, on the one hand, nor like that of traditional Artificial Intelligence on the other.

Acknowledgments

References

Abbott, V., Black, J. and Smith, E. (1985). The representation of scripts in memory. *Journal of Memory and Language*, 24, 179–199.

Barlow, H. B. (1985). Cerebral cortex as a model builder. In D. Rose and V. Dobson (eds), *Models of the Visual Cortex*. John Wiley.

Barsalou, L. and Sewell, D. (1985). Contrasting the representation of scripts and categories. *Journal of Memory and Language*, 24, 646–665.

Braitenberg, V. (1978). Cortical architectonics: general and areal. In M. A. Brazier and H. Petsche (eds), *Architectonics of the Cerebral Cortex*. New York: Raven Press.

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.

Calvin, W. (1987). *The River that Flows Uphill: A Voyage from the Big Bang to the Big Brain*. New York: Macmillan.

Craik, K. J. W. (1943). *The Nature of Explanation*. London: Cambridge Unversity Press.

Eccles, J. C. (1957). *The Physiology of Nerve Cells*. Baltimore: The Johns Hopkins Press.

Edelman, G. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, 95, 307–317.

Feldman, J. A. (1988). Connectionist representation of concepts. In D. Waltz and J. A. Feldman (eds), *Connectionist Models and their Implications: Readings from Cognitive Science*. Norwood, NJ: Ablex.

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

French, R. and Weaver, M. (1987). The role of categories in the generation of counterfactuals: a connectionist interpretation. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates.

Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychological Bulletin*, 38, 781–817.

Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11.

Hebb, D. O. (1949). *The Organization of Behavior*. John Wiley.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books.

Hume, D. (1737/1975). Of the idea of necessary connexions. In *An Enquiry Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press.

Johnson-Laird, P. N. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society*, London 295B, 353–374.

Jonides, J. (1980). Towards a model of the minds eye movement. *Canadian Journal of Psychology*, 34, 103.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

Kaplan, S. (1970). The role of location processing in the perception of the environment. In *edra two* (Proceedings of the second annual environmental design research association conference, Pittsburgh), J. Archea and C. Eastman (eds).

Kaplan, S. (1973). Cognitive maps in perception and thought. In R. M. Downs and D. Stea (eds), *Image and Environment*, Chicago: Aldine.

Kaplan, S. (1976). Adaptation, structure, and knowledge. In G. T. Moore and R. G. Golledge (eds), *Environmental Knowing*. Stroudsburg, PA: Dowden, Hutchinson, and Ross.

Kaplan, S. (in press). Beyond rationality: clarity based decision making. In G. Evans and T. Garling, (eds), *Environmental Cognition and Assessment*.

Kaplan, S. and Kaplan, R. (1982). *Cognition and Environment*. Praeger.

Kinsbourne, M. (1982). Hemispheric specialization and the growth of human understanding. *American Psychologist*, 37(4), 411–420.

Kleinmuntz, B. (1968). The processing of clinical information by man and machine. In B. Kleinmuntz (ed.), *Formal Representation of Human Judgement*. New York: Wiley.

Kosslyn, S. (1987). Seeing and imaging in the cerebral hemispheres: a computational approach. *Psychological Review*, 94(2), 148–175.

Kruglanski, A. and Ajzen, I. (1983). Bias and error in human judgement. *European Journal of Social Psychology*, 13, 1–44.

Lachman, J. L. and Lachman, R. (1979). Theories of memory orgnization and human evolution, In C. R. Puff (ed.), *Memory Organization and Structure*. Academic Press.

Lachter, J. and Bever, T. (1988). The relationship between linguistic structure and associative

theories of language learning – A constructive critique of some connectionist teaching models. *Cognition*, 28, 195–247.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. In *Society of Experimental Biology Symposium No. 4: Psychological Mechanisms in Animal Behaviour* (pp. 478–505). London: Cambridge University Press.

Lindsay, R. K. (1988). Can this treatment raise the dead? *Behavioral and Brain Sciences*, 10(4).

Luria, A. R. (1973). *The Working Brain*. Penguin Books.

Lynch, J. C., Mountcastle, V. B., Talbot, W. H. and Yin, T. C. T. (1977). Parietal lobe mechanisms for directed visual attention, *Journal of Neurophysiology*, 40(2), 362–389.

Macphail, E. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences*, 10(4).

Margolis, H. (1979). *Patterns, Thinking, and Cognition*. Chicago, IL: University of Chicago Press.

Meyer, D. and Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology: General*, 106, 227–234.

Miller, G., Galanter, E., and Pribram, K. (1960). *Plans and the Structure of Behavior*. New York: Henry Holt.

Miller, N. (1963). Some reflections on the law of effect produce a new alternative to drive reduction. *Nebraska Symposium on Motivation 1963*. Lincoln: University of Nebraska Press.

Milner, P. M. (1957). The cell assembly: Mark II. *Psychological Review*, 64.

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

Nisbett, R. and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. (1980). Cognitive engineering and education. In D. Tuma and F. Reif (eds), *Problem Solving and Education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Oakesford, M., Chater, N. and Stenning, K. Connectionism, Classical Cognitive Science and Experimental Psychology. Submitted to *AI and Society*.

Palm, G. (1982). *Neural Assemblies: An Alternative Approach to Artificial Intelligence*. Springer-Verlag, New York.

Posner, M. I. (1973). Search strategies and problem solving. In *Cognition: An Introduction*. Scott Foresman.

Prince, A. and Pinker, S. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28.

Pylyshyn, Z. W. (1980). Cognition and computation: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(1), 154–169.

Rochester, N., Holland, J. H., Haibt, and Duda, W. L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, IT-2.

Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 7, 573–605.

Rosenberg, C. (1987). Revealing the structure of NETtalk's internal representations. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates.

Rumelhart, D., and McClelland, J. (1987). On learning the past tenses of English verbs. In J. McClelland and Rumelhart, D. (ed), *Parallel Distributed Processing, Vol. 2*. Bradford.

Sejnowski, T. and Rosenberg, C. (1986). NETtalk: A parallel network that learns to read aloud (Technical Report 86/101). Baltimore, Johns Hopkins University, Department of Electrical Engineering and Computer Science.

Shepard, R. (1975). Form, formation and transforation of representation. In R. Solso (ed.). *Information Processing and Cognition*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, E., and Medin, D. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smolensky, P. (1988). The constituent structure of mental states: a reply to Fodor and Pylyshyn. Technical report CU–CS–394–88, Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder.

Stea, D. (1978). Environment, perception and cognition: towards a model for mental maps. In S. Kaplan and R. Kaplan (eds) *Humanscape: Environments for People*. Belmont, CA: Duxbury.

Tolman, E. C. and Brunswick, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, 42, 43–77.

Tolman, E. C. (1948). Cognitive maps in rats and man. *Psychological Review*, 55, 189.

Voss, J. F. (1969). Associative learning and thought. In J. Voss (ed.) *Approaches to Thought*, Columbus, OH: Charles E. Merrill.

Voss, J. F., Tyler, S. W. and Yengo, L. A. (1983). Individual differences in the solving of social

science problems. In R. F. Dillon and R. R. Schmeck (eds), *Individual Differences in Cognition*. New York: Academic Press.

Wason, P. C. and Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, Mass.: Harvard University Press.

Weinert, F. and Kluwe, R. (eds) (1987). *Metacognition, Motivation, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, R. J. and Zipser, D. (1988). A learning algorithm for continually running, fully-recurrent neural networks. Technical Report no. 8805, Institute of Cognitive Science, University of California at San Diego, La Jolla, CA.

Wise, S. P. and Desimone, R. (1988). Behavioral neurophysiology: insights into seeing and grasping, *Science*, 242, 736–741.

Correspondence and offprint requests to: Stephen Kaplan, Department of Psychology and Department of Electrical Engineering and Computer Science, 3418 Mason Hall, The University of Michigan, Ann Arbor, MI 48109, USA

Tel: (313) 764 0426, email: Stephen__Kaplan@*um.umich.edu*