CONNECTIONISM: IS IT A PARADIGM SHIFT FOR PSYCHOLOGY?

Techinical Report AIP - 24

Walter Schneider

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

29 September 1987

This research was supported by the Computer Sciences Division, Office of Naval Research and DARPA under Contract Number N00014-86-K-0678 and N00014-86-0107. I wish to acknowledge the many rewarding interactions I have had with Jay McClelland and Geoffrey Hinton on the topic of connectionism.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 24	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research (Code 1133)
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Same as Monitoring Organization 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS p400005ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO: N/A; TASK NO: N/A; WORK UNIT ACCESSION NO: N/A

11. TITLE (Include Security Classification)
Connectionism: Is it a Paradigm Shift for Psychology?

12. PERSONAL AUTHOR(S)
Walter Schneider

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	87 September 29	11

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Connectionism, Cognitive Psychology

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Connectionism is a method of modeling cognition as the interaction of neuron-like units. Connectionism has received a great deal of interest and may represent a paradigm shift for psychology. The nature of a paradigm shift (Kuhn, 1970) is reviewed with respect to connectionism. The reader is provided an overview on connectionism including: an introduction to connectionist modeling, new issues it emphasizes, a brief history, its developing sociopolitical impact, theoretical impact, and empirical impact. Cautions, concerns, and enthusiasm for connectionism are expressed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b. TELEPHONE (Include Area Code) (202) 696-4302	22c. OFFICE SYMBOL N00014

Behavior Research Methods, Instruments, & Computers
1987, 19 (2), 73-83

SESSION I
PRESIDENTIAL ADDRESS

Connectionism: Is it a paradigm shift for psychology?

WALTER SCHNEIDER
University of Pittsburgh, Pittsburgh, Pennsylvania

Connectionism is a method of modeling cognition as the interaction of neuron-like units. Connectionism has received a gread deal of interest and may represent a paradigm shift for psychology. The nature of a paradigm shift (Kuhn, 1970) is reviewed with respect to connectionism. The reader is provided an overview on connectionism including: an introduction to connectionist modeling, new issues it emphasizes, a brief history, its developing sociopolitical impact, theoretical impact, and empirical impact. Cautions, concerns, and enthusiasm for connectionism are expressed.

In recent years there has been an explosive interest in modeling cognition within a connectionist framework. The connectionist framework assumes that cognition is carried out via the mutual interaction of neuron-like elements. The theoretical interest in this approach probably represents the most dramatic shift in theoretical orientation in psychology in the last 20 years. This modeling is still in its infancy. We are currently in a period of exciting development. In this presidential address, I review some of the basics of connectionist modeling and describe the reasons for the enthusiasm and some reasons for caution. I also encourage the reader to try to decide for himself/herself whether or not this represents a paradigm shift in the sense of Kuhn (1970).

Throughout the history of psychology, we have generally tried to describe the brain in terms of the most complex systems we understand. In this century the brain has been described in terms of a telephone network, a homeostatic system, a computer system, a semantic net, and a production system. Connectionism is different: it seeks to model cognition in terms of something we do not understand, that is, how the brain operates. It utilizes very simplisitic features of the brain's physiology to attempt to model cognitive processes. Connectionism examines computation based on the assumption of many parallel processing elements. Each element combines simple analog inputs weighted by the strength of the connection to produce analog or digital outputs. Connectionism does not incorporate either the microstructure (e.g., differential polarization, depending on whether the synapse contacts the cell body or the dendrite) or macrostructure (e.g., very specific neuroanatomical connections between regions of the cortex) of neurophysiology (see Sejnowski, 1986). However, the simplifications do make the models tractable and allow us to begin looking at what neural-like systems could compute. As a result of dissatisfaction with previous modeling frameworks and an availability of computer resources, a number of researchers have begun a movement toward modeling connectionist systems.

CHARACTERISTICS OF A PARADIGM SHIFT

It is useful to review some of the characteristics of a paradigm shift according to Kuhn (1970). Four characteristics of a paradigm shift seem to be present in the current movement toward connectionism. Kuhn commented that "all crises begin with a blurring of the paradigm and a consequent loosening of the rules for normal research" (p. 84). This loosening typically occurs partially because few practitioners agree on what the paradigm is. In the 1970s there was a clear movement away from box models of information processing to a variety of representations (e.g., levels of processing, schemata, semantic networks, and production systems). One of the examples of this loosening is that a number of psychologists are now studying learning in computer models rather than explicitly examining learning in humans. Kuhn commented that anomalies appear that do not fit the traditional view (pp. 82-91). In psychology, due to our relatively weak theories, there are many phenomena that we poorly predict. Two phenomena that are particularly important from the connectionist perspective are our abilities to learn without instruction and to perform procedural tasks very well even when we are unable to specify the rules of that perfor-

I wish to acknowledge the many rewarding interactions I have had with Jay McClelland and Geoffrey Hinton on the topic of connectionism. My own research on simulation modeling is supported by Contract No. N00014-86-0107 from the Office of Naval Research. Reprint requests should be addressed to Walter Schneider, 517 Learning Research & Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260.

mance. The difficulty of obtaining knowledge from experts to build expert systems illustrates the problems of rule-based descriptions.

Kuhn (1970) suggested that a new paradigm must provide the hope that it is possible to march forward (p. 158). The connectionist framework suggests that we might be able to connect the computational, cognitive, and physiological levels of analysis and to do so with a conceptually very simple system. During a paradigm shift "communication across the revolutionary divide is inevitably partial" (p. 149). Connectionism is introducing new vocabulary (e.g., vectors, weight spaces), new mathematics (e.g., eigenvectors, gradient descent), and even new rules of evidence in psychology (e.g., posing simulation experiments about small-scale learning systems to illustrate what can be learned by such systems). Finally, Kuhn stated that "during the transition period there will be a large but never complete overlap between the problems that can be solved by the old and the new paradigm" (p. 85). For example, connectionism and production systems both examine learning. However, connectionism focuses on slow learning, such as learning the correspondence between text and speech, which may require 40,000 trials of training (e.g., Sejnowski & Rosenberg, 1986). Production system learning typically examines learning that occurs in under 10 trials (e.g., J. R. Anderson, 1983).

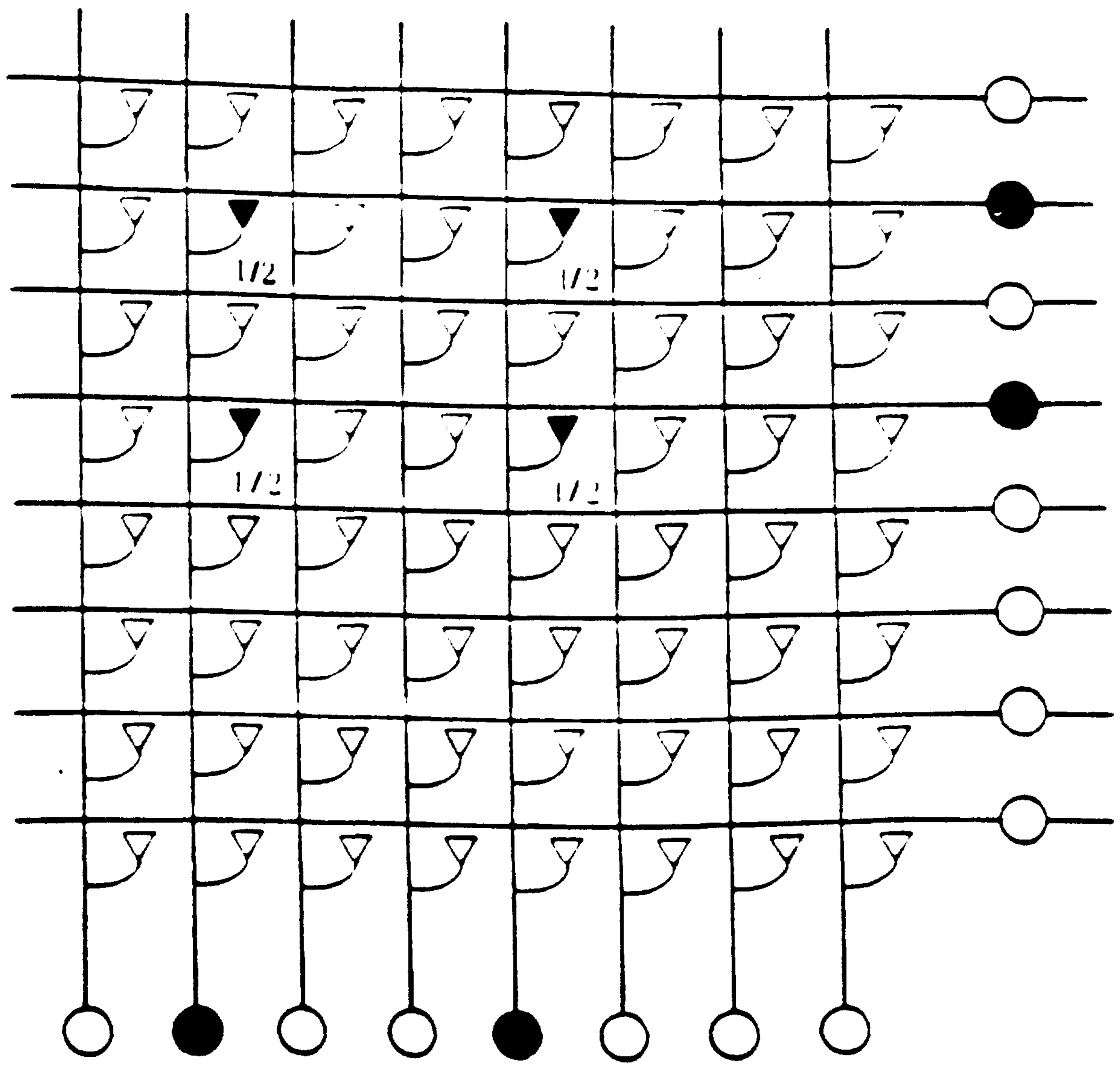

Figure 1. A connectionist association matrix. The input units are on the bottom, the output units on the right. The triangles represent connections from the input to the output. The filled circles represent the active units. Learning involves changing the strength of the input units to the output units. The filled triangles illustrate which connections would change so that the input would evoke the output. The figure is adapted from Figure 1 in "Resource Requirements of Standard and Programmable Nets" by J. L. McClelland, 1986, in D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Volume 1: Foundations*, p. 462. Copyright 1986 by MIT Press. Adapted by permission.

DEFINING FEATURES OF CONNECTIONIST MODELS

Four defining features are common to all connectionist models. First, processing is assumed to occur in populations of simple elements. The letter H, for example, may be encoded as a set of eight elements that have binary values for features, such as vertically symmetric, horizontally symmetric, diagonally symmetric, not rounded, not diagonal, not closed, and without descender. Although some information may be encoded by a single element being on, most information is coded by a set of elements being on or a vector of activation.

The second, and perhaps prototypical, characteristic is that all knowledge is stored in the connectionist weights between the elements. Knowledge is stored in the associations or strength of connections between neural-like elements (see Figure 1). The knowledge is stored in a small number of association matrices that represent the addition of all the stimulus response patterns the system has learned. This results in making the knowledge very context sensitive. For example, it may be more difficult to learn the past tense of *go* as being *went* because for most words the past tense of words is formed by adding *ed* (see Rumelhart & McClelland, 1986a).

The third characteristic is that all the units perform a simple combination of their inputs (e.g., addition or multiplication) and perform a simple nonlinear transformation on those inputs (e.g., a logistic function). There is generally no complex matching of a particular set of inputs to a unit to some internal pattern (e.g., as might occur in a symbol-processing-based comparison). Rather, a unit generally simply adds or multiplies all the inputs. The nonlinear transformation is sometimes represented as a simple saturation effect (e.g., a neuron can fire at a frequency of no less than 0 and no more than 1,000 times per second). This nonlinearity is critical in that it gives the models the ability to categorize information (J. A. Anderson & Mozer, 1981).

The fourth characteristic is that learning occurs via simple learning rules that are based on local information available within the unit. Learning involves modifying the connections to enable a later input pattern to evoke a new output pattern. There are a variety of learning rules that have been employed (see Rumelhart & McClelland, 1986b). In order to associate an input to an output, the weights between the input and output units are modified so that the input unit will evoke the output. Figure 1 shows a simple illustration of the delta learning rule. If two units were on in the input layer and two units were on in the output layer, the connection strength between the input and output units would be increased to a level of a desired output of one divided by the number of input neurons that were on. This results in the input pattern becoming able to evoke the output pattern. In order to reduce the interference between different input patterns in the same association network, a variety of more sophisticated learning rules (e.g., delta rule, Boltzmann learning, back propagation algorithm; see Rumelhart, Hinton, & Williams, 1986) are utilized.

EXAMPLE OF CONNECTIONIST LEARNING

There are six basic steps in conducting a connectionist simulation. First, the input and output units and codes for the model must be specified. Second, the connection architecture specifying the number of units at the input, output, and any intermediate layers of processing must be established. Third, the initial weights must be set to small random values. Fourth, the input and the desired output must be presented for all the input and output relations to be learned. Fifth, some learning rule must be applied such that the weights are updated so that the input comes to activate the output. The simulation may present the presentation and learning steps hundreds of thousands of times. Sixth, diagnostic experiments (e.g., presenting degraded stimuli, cutting out connections, examining transfer to related patterns) must be run to determine the robustness and generalizability of the knowledge.

Probably the flashiest demonstration of connectionist learning is embodied in NET-TALK by Sejnowski and Rosenberg (1986) (see Figure 2). They taught a network to learn to associate English text to the appropriate English phonology. There were seven groups of letter positions of visual input. Each position could be encoded as one of 29 characters including punctuation. There were 26 output feature units coding one of 53 potential phonemes. The intermediate or hidden units recoded the input to produce the desired output. The model was presented successive passes over the text and the phonology of a corpus of 1,024 words of continuous informal speech produced by a child. After 10,000 presentations of words, the network was about 85% accurate at specifying the phonemes for the text input. An accuracy of 90% was reached by 20,000 trials and of 95% by 50,000 trials. The demonstration is particularly memorable because one can listen to the network speak. The output of the network controls a synthetic speech production system. During the initial learning, the system babbles, continuously outputting a few vowels. It gradually learns to distinguish between vowels and consonants, and then it learns to identify the space as a pause. The system begins to babble in pseudospeech form and gradually acquires some words. After 40,000 trials, it produces words that sound intuitively like those you might expect to hear from a 2-year-old child. This demonstration is very intriguing, and the auditory tape produced by the network has been played many times, including once on network television on the "Today Show."

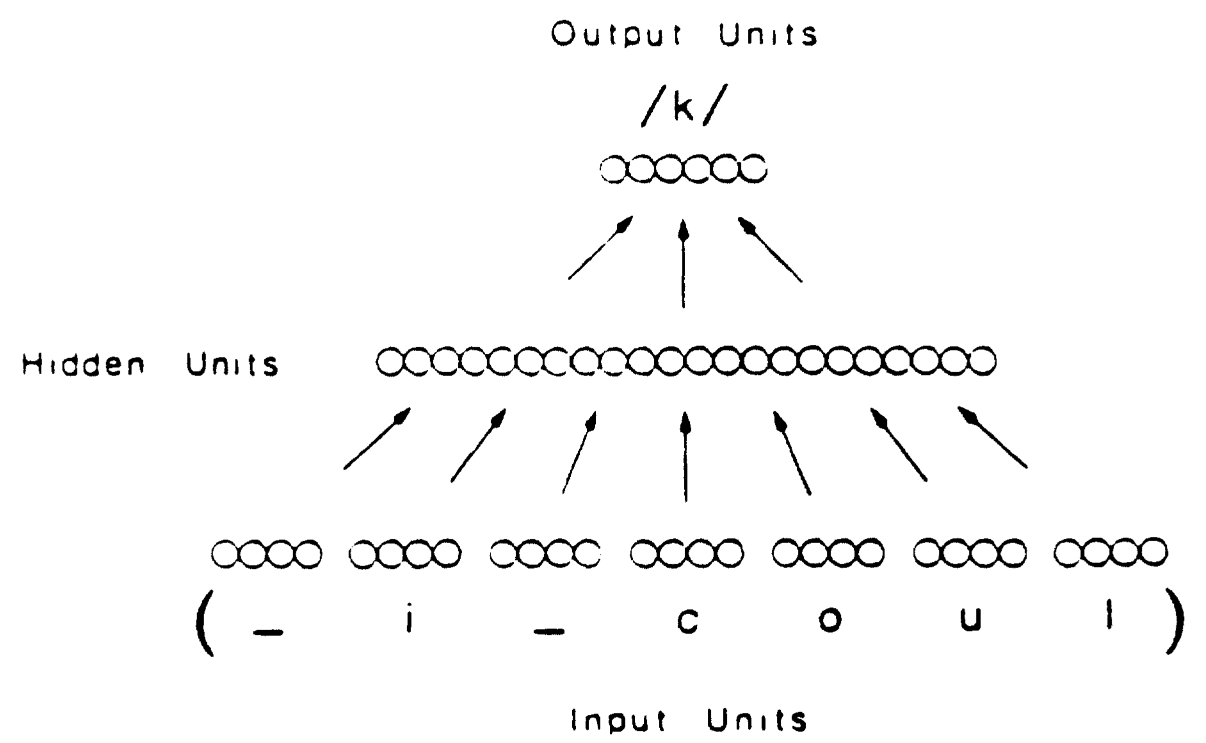

Figure 2. Schematic drawing of the Sejnowski and Rosenberg (1986) NET-TALK connection architecture. Input units are shown on the bottom of the pyramid, with seven groups for sequential letter positions. Each hidden unit in the intermediate layer receives inputs from all of the input units on the bottom layer, and in turn sends its outputs to all 26 phonemic feature units in the output layer. An example of an input string of letters is shown below the input groups, and the correct output phoneme for the middle letter is shown above the output layer. The network was presented letter strings and phonemic patterns. The connection weights were altered using back propagation. From T. J. Sejnowski and C.R. Rosenberg, 1986, *NETtalk: A Parallel Network that Learns to Read Aloud* (Tech. Rep. No. JHU/EECS-86/01), The Johns Hopkins University Electrical Engineering and Computer Science, Baltimore, MD.

With a working connectionist model in hand, there are a variety of experiments that can be performed. First, one can look at the type of units developed to perform the task. This is done by examining the input and output weights for each of the units. The units each specialize in performing some complex functional transformation of the input to the output. It is generally very difficult to interpret the form of the units. The units operate in very high-dimensional spaces (e.g., 80 dimensions). Examining any one unit in isolation provides one with little information about what the network is doing as a whole. The information is distributed across all of the units in the network. After the network has learned to map a particular input to an output, one can examine how well this learning generalizes to novel words. NET-TALK reproduced correctly 78% of the novel words it was presented. One can also examine how the network reacts to damage to the network. These systems are typically quite robust to substantial amounts of damage in the network (e.g., J. A. Anderson, 1983). NET-TALK illustrated that relearning after damage to the network can be substantially faster (i.e., 10 times faster) than the original learning. One can also explore such issues as how learning changes as a function of the number of units in the intermediate layers.

A PARADIGM SHIFT EXPOSES NEW ISSUES

A paradigm shift emphasizes new issues. These are often issues that existed in the field before but now are brought to center stage for close examination. Four issues are particularly important in the connectionist paradigm. The issue of representation, the hidden units problem and learning rules, the problem of sequencing, and the nature of teaching.

The representational issue involves coding information so that connectionist networks can perform nontrivial information processing tasks. For example, if one wants a model to perceive words exhibiting behaviors that humans produce, should the model have levels for visual features, letters, and word units (e.g., see McClelland & Rumelhart, 1981)? What are the semantic features of nouns (McClelland & Kawamoto, 1986)? How are family relationships coded in a network (Hinton, 1986)? In order to produce a workable model, people have to become very explicit as to what information is stored in a network. Rumelhart and McClelland (1986a) were unable to have their simulation accurately associate word phonemes to the phonemes for the past tense of words using a number of coding schemes. They then tried coding words in terms of Wickelphones (a scheme proposed by Wickelgren [1969] to code a phoneme in the context of its preceding and following phoneme). With this coding scheme the networks could learn to associate words with the past tense sound of the words. Producing representations that are learnable in realistic time periods provides a serious constraint on connectionist models. These constraints allow the use of learnability constraints to evaluate representations.

Connectionism has given considerable emphasis to the "hidden unit problem" (Hinton & Sejnowski, 1986). In order to learn complex responses to a given input pattern, one cannot simply connect the inputs to the output units. If one directly connects the input units to the output units, only first-order relationships can be learned. For example, if two inputs are connected to one output, the network can learn to perform either an AND or an OR operation. However, it cannot learn to perform an exclusive XOR operation (i.e, "on" if either of the inputs are on; "off" either if both of the inputs are off or if both of the inputs are "on"). A network cannot learn such second-order information with only pair-wise weights between the visible units (i.e., the input and output units). In order to learn such input/output relationships, a set of hidden units are needed that receive connections from the input unit and make connections to the output units. However, the hidden units themselves are not set directly by either the input or the output. Changes in the connection strength in the hidden units reorganize the input pattern to allow the learning of more complex input/output patterns (Rumelhart, Hinton, & Williams, 1986). Algorithms that enable hidden unit learning develop truly emergent properties. For example, networks with hidden units can solve the XOR problem (Ackley, Hinton, & Sejnowski, 1985). NET-TALK reached only an 80% accuracy in a network without hidden units, whereas it reached a 95% accuracy with hidden units. The study of the hidden unit problem has emphasized the need to understand the nature of higher order similarity. Human learning is very much influenced by similarity. Traditional approaches to learning have had relatively poor techniques for interpreting and predicting these similarity effects.

The third issue emphasized in connectionist simulations is the problem sequencing. For example, should training proceed by first showing the prototypes of a category and then showing the more distant exemplars? As networks are presented examples, they perform a search through a weight space (i.e., the strengths of all the connections), trying to come up with the best combination of weights. Depending upon whether practice is distributed or massed, differential learning is observed that looks similar to that seen in humans (see Rosenberg & Sejnowski, 1986). Connectionism emphasizes learning rules that can rapidly modify weights so that the hidden units can perform complex computations (e.g., Boltzmann learning, back propagation; see Rumelhart, Hinton, & Williams, 1986).

The fourth issue in connectionism is an explicit concern for various levels of teaching. Connectionist networks can learn in one of three types of learning or supervision environments. The first class is *supervised learning*, in which a teacher explicitly indicates to the network what the correct output state is for any input state. In this sense, the teacher is a supervisor. The network then compares the output produced by the input to the desired output and uses that difference in activation to modify the weights in the network. The NET-TALK example is an instance of supervised learning. Supervised learning is slow initially, but the network can very quickly acquire new associations that are similar to previous associations.

The second class of learning involves a yes-no teacher and is referred to as *reinforcement learning*. In such a situation the teacher provides the learner feedback only at the end of a trial, after the student has executed many operations. Barto and Anandan (1985) taught a connectionist network to perform a pole-balancing operation on a moving cart. The network would push the stick left or right, trying to balance it on the cart as long as possible. Eventually, after many stick movements, the cart would run into a barrier on the left or right side. This running into the barrier was the only feedback the network received. The network then had to learn when to push the pole to the left or right to try to balance it so that the cart would stay between the two barriers. The stick might be moved a hundred times before the cart would hit one of the barriers. The system learned to perform this task by dividing the learning into two components. The controller network controlled the stick and performed operations similar to supervised learning. However, the supervision was provided by a second teacher network. This network used the input from the controller to try to predict whether or not a "yes" or a "no" would come from the teacher (i.e., whether it would hit a barrier). The teacher network developed the ability to predict error signals that the supervised learning teacher would provide during the time preceding the "yes/no" reinforcement. The teacher network used this information to give feedback to the controller network. The controller network then learned via supervised-like learning procedures and eventually acquired the skill. It should be noted that learning under this procedure is far slower than learning via supervised learning procedures.

The third class of learning is *unsupervised learning*, or learning without any teacher at all. Under this type of learning, the system tries to predict its own behavior through a small number of hidden units. For example, Elman and Zipser (1987) used unsupervised learning to have a network learn the basic features of speech phonetic perception. In their model they used 50 input units for portions of the speech spectrogram, 20 hidden units, and 50 output units that predicted the speech spectrogram. The input pattern activated the hidden units, and the hidden units activated output units that paralleled the input units. The network was able to compare the input to what it produced from that coded version of the input. Since the hidden unit level contained far fewer units than the input or output level, the hidden units had to develop some type of generalized scheme for coding the information. The hidden units captured the major higher order invariances of the input. Elman and Zipser (1987) presented the acoustic stimulus, "this is the voice of the neural network," to the network 100,000 times. Then the hidden units captured sufficient features of the input so that the network could reproduce the speech quite intelligibly. More importantly, the network captured generalizations of the inputs. The hidden units were, in essence, encoding the stimulus in phoneme-like feature codes that could be used for higher levels of processing. Using unsupervised learning, a network can develop representations of higher-order invariances of the external world as a result of mere exposure. This type of unsupervised learning suggests how the Suzuki method of teaching violin might be effective. A student who repeatedly hears certain acoustic patterns learns to encode those features of the pattern. This encoding can be used later to verify whether the student can produce the desired acoustic code. More generally, this unsupervised learning provides an interpretation of how listening to speech might help a child learn the phonemes of the target language in the absence of corrective feedback.

BRIEF HISTORY OF CONNECTIONISM

In the short history of connectionism in psychology, it has already had a birth, a death, and a rebirth (see Rumelhart & McClelland, 1986b, for detailed account). In the late 1950s the perceptron was a basic connectionist network with no hidden units. This system was proposed as a neurally feasible mechanism that could accomplish complex learning (Rosenblatt, 1962). In 1969, Minsky and Papert provided a very severe and influential critique that suggested that the study of perceptrons would be "sterile" because it could not deal with the hidden unit problem. The field was fairly dormant for about 10 years. By 1981 there was a substantial rebirth of interest in perceptron-type models as illustrated by the publication of the book *Parallel Models of Associative Memory* by Hinton and J. A. Anderson (1981). By 1985 the Minsky and Papert critique was finally confronted and overcome with the solution of the hidden unit problem by Ackley et al. (1985). Shortly thereafter, Rumelhart, Hinton, and Williams (1986) developed the back propagation algorithm that allowed very rapid computer simulation of learning for networks with hidden units. With NET-TALK, Sejnowski and Rosenberg (1986) provided a very imaginative and enthusiastic demonstration of connectionist learning processes. In 1986 Rumelhart and McClelland and McClelland and Rumelhart provided a two-volume textbook entitled *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. These volumes provide a 1,158-page compendium of the techniques and simulations of connectionism. The books provide a wealth of new connectionist modeling simulations and concepts. The volumes are likely to be classics and are the basis for many courses in connectionism throughout the country.

SOCIOPOLITICAL IMPACT OF THE SHIFT

A paradigm shift has a substantial social and political impact on a field. Connectionism is certainly having such an impact. First, there is a great deal of excitement and interest in the topic. Many young and older researchers are exploring such modeling. Connectionists seminars are probably occurring in a hundred universities in the country this year. Established researchers, such as Walter Kinsch, Earl Hunt, Danny Kahneman, and Gordon Bower, are examining or applying connectionist models to their work. The sales of the *Parallel Distributed Processing* books have been phenomenal. The books literally sold out (6,000 copies) before they went to press. One wonders if psychology has ever before had a two-volume advanced textbook sell-out. The rapid growth of connectionist talks at the Cognitive Science Society meetings illustrates this exciting interest: in the years 1984, 1985, and 1986, the percentage of connectionist talks were 17%, 23%, and 31%, respectively. In a period of about 5 years, connectionism went from being nearly nonexistent to being one third of the program of the Cognitive Science Society.

Granting agencies have also shifted toward connectionism. The Sloan Foundation, the National Science Foundation, the Office of Naval Research, the Defense Advanced Research Project Agency, and the Air Force Office of Scientific Research all have initiated programs to fund this type of modeling. This modeling has caught the interest of basic researchers who wish to understand cognition and biological computing, as well as of applied researchers who want to build better weapon systems. Note this shift in cognitive science has in some cases reduced funds available for experimental research. Thus there is a shift in the research base for the future.

In the summer of 1986 there was a connectionist summer camp. Under Sloan Foundation sponsorship, Sejnowski, Hinton, and Touretzky brought together 50 graduate students for an 11-day workshop on connectionism. The goal was explicitly to seed the world with connectionists. The workshop brought these researchers together so they could exchange techniques and develop substantial enthusiasm for changing the field.

More important than changing the social climate, connectionism is altering the conceptual environment. McClelland, for example, describes sentence processing as not being grammar processing, but rather as being the unitization of a set of clues to interpret meaning. Rumelhart describes "representations as being built not specified." The ability to use large quantities of information in an interactive manner allows conceptualization of processing in a manner very different from that of serial computers.

The impact of connectionism is likely to go well beyond the psychological laboratory. Hammerstrom (1986), a computer architect, predicts that "it will be possible within 5-10 years to build a silicon-based system that emulates a network of a billion connections between millions of nodes," and these systems "will be relatively cheap" (approximately $300 for production costs) and compact (size of a floppy disk), simulating neural systems at roughly two orders of magnitude faster than real time. Think of the implications, perhaps in 20 years, of having the processing capacity of our speech processing available for a $300 device that can be connected to a personal computer. If these learning systems can perform perceptual and learning activities that we currently associate with humans, this connectionism movement will cause a second computer revolution that would be more significant than the first.

THEORETICAL IMPACT

The theoretical impact of connectionism on psychology is strong and likely to be great. Connectionism is making theories of learning much more explicit. For these models one must describe the number of elements at each level, the internal codes, the problem sequencing rules, and the learning algorithms.

Connectionism allows new types of studies. Most connectionist modelers are examining the psychology of nonhuman intelligence systems. The typical procedure is to build a network-type robot to see what it learns on its own. This is an engineering approach with simulation providing existence proofs. It should be noted that this method of existence proofs has been very productive in computer science by developing a basis of algorithms and procedures. It may help the psychology of cognition to become a more cumulative endeavor.

Connectionism has introduced a variety of new (improved) concepts and language. We can now discuss representations in terms of vector spaces. Learning is described as a method of gradient descent or learning by approximation. We can categorize the type of supervision of the learning process and how the problems should be sequenced to maximize learning. All of these issues can now be tested with simulations providing quantitative data.

Connectionism has provided a new emphasis to a number of psychological phenomena. McClelland and Rumelhart (1981) emphasized the importance of top-down influences in the word superiority effect. Ackley, Hinton, and Sejnowski (1985) described mechanisms that enable unsupervised learning to acquire complex relationships. Hinton and Plaut (in press) illustrated how relearning can be much faster than original learning and can even transfer to material that was not explicitly taught. For example, if one has not used a foreign language for many years, learning to use a subset of the words of that language can show substantial transfer to words that were not explicitly relearned. Hinton refers to this process as compensating for the defocusing of memory across time. Hinton and Nowlan (in press) recently described how a learning mechanism can greatly speed evolution. In this system, genes can either be in one of two states or be in a modifiable/learnable state. He shows that with learnable states, individual learning trials can be substituted for generations. Given that learning trials are very cheap compared with spawning a new generation, this learning mechanism can greatly speed evolution.

CONNECTIONIST REFORMULATION OF PSYCHOLOGICAL CONCEPTS

There are three formulations of psychological concepts provided by connectionism that I find particularly interesting and exciting. All of these concepts existed before connectionism, but the concepts have become more concrete and elegant within the connectionist framework.

The concept of a semantic network can be recast within a connectionist framework. In a semantic network one typically has "Is-A" links between nodes in a network. For example, in a semantic network of family relationships, one might have the names of family members connected with "Is-A son," "Is-A father," "Is-A daughter," and so forth. One of the problems of the semantic network is that if the network is taught only a subset of the links, it must use some complex strategies to find new relationships. For example, if the system is taught that Jim is the son of Jack and that Sue is the daughter of Jack, the system does not directly generalize that Jim and Sue are siblings. This can be done with complex postretrieval processing where various alternative link combinations are examined to infer whether the sibling relationship holds. Hinton (1986) taught a connectionist network to learn family relationships. The system was required to learn 100 relationships among 24 names from two families. There were 24 input names, 12 family relationships, and 24 output names. In addition there were 12 hidden units representing the input family, 12 hidden units representing the output family, 6 hidden units for the relationship, and 12 central representational units. The system was taught 100 of the 104 instances of relationships (e.g., father, mother, husband, wife, son, daughter, etc.). The 12-name hidden units learned to code relationships. The hidden units recoded input names in terms of their generation level and family type. Note that this recoding rule was developed by the network as a result of presenting family relationships and the network applying a simple (i.e., back propagation) learning rule to change the weights of the hidden units. The hidden units encode individual names in terms of family relationships (e.g., generation, sex). If the system is [illegible] Jim is the son of Jack and that Sue is the daughter of Jack, the system will infer (via generation and relational coding) that Jim is the brother of Sue. This is done without any complex postprocessing, but rather is a side effect of building an internal representation for the family codes. This kind of coding might explain why a parent may make the verbal slip of calling a child by the name of one of his or her siblings. Connectionism provides a very simple interpretation of these phenomena and how both the encoding and retrieval processes can be accomplished with a simple parallel distributed operation.

Connectionism enables recasting schemata within the concrete representational framework. The concept of schemata has been around for a long time and is felt by some researchers to be a major building block of recognition (Rumelhart, 1980). Generally the representations of schemata have been vague specifications of a grouping of elements that co-occur in some expected fashion. In the connectionist framework, schema theory can have an explicit form that can predict the interrelationships of objects (Rumelhart, Smolensky, McClelland, & Hinton, 1986). The elements of the schema can be represented as individual units in a connectionist network. The strengths of the connection between the units are determined by the co-occurrence frequency of the various objects of the schema. For example, Rumelhart had subjects list the objects that one would typically find in a living room, bathroom, study, etc. The strengths of connections between the elements were determined by the co-occurrence frequency of the elements. Accordingly, *bookshelf* and *desk* would have a very strong co-occurrence frequency, whereas *bookshelf* and *oven* would not. The connections between the units for *bookshelf* and *desk* would have a strong weight; *bookshelf* and *oven* would not. In a simulation, two of the 40 units would be activated, and the activation of the others would be measured. This activation represented the filling in of the schema elements. For example, the activation of *desk* and *ceiling* would activate the terms *computer*, *books*, *bookshelf*, *typewriter*, *doors*, and *walls*. In contrast, activating *bathtub* and *ceiling* would result in the activation of *scale*, *toilet*, *very small*, and *walls*. If such unexpected combinations as *sofa*, *bed*, and *ceiling* were activated, novel configurations of rooms would be activated including *television*, *dresser*, *drapes*, *fireplace*, *books*, and *large*. This connectionist network illustrates how schemata can be built up and can fill in missing information, as well as misinterpret information, to make it more consistent with the current schema. All of the current operations occur through the simple mechanism of the parallel distributed activation of the elements that might occur in a room.

The third example of connectionism's recasting of a vague concept into an explicit form is one of my own. In 1977 Shiffrin and Schneider described a dual processing model in which the two forms of processing were called automatic and controlled. Figure 3A illustrates the original figure. Automatic processing was viewed as fast, parallel, and fairly effortless. In contrast, controlled processing was viewed as a slow, typically serial, and effortful form of processing. At that time we could not provide a mechanism of these two qualitatively different forms of processing. Recently, Schneider and Mumme (1987) recast the concept of automatic and controlled processing within a connectionist architecture (see Figure 3B). Controlled processing involves an external source that modulates the output of all of the elements from a module. Automatic processing involves a local circuit (through the priority report cell), which enables the output of a module in the absence of an external attentional input. Within each module there is a connectionist association of the input patterns to a priority tag for that message. If that message is of high enough priority, the message is automatically transmitted in the absence of controlled processing input. The priority mechanism produces the four phenomena of automatic processing as emergent properties. That is, as automatic processing develops, performance becomes fast, effortless, and difficult to control, and it results in reduced ability to modify memory (see Schneider & Mumme, 1987; Schneider & Detweiler, 1987). The connectionist model predicts how performance shifts from a serial to a parallel processor as practice continues in a consistent search paradigm. The simulation also illustrates that even though the mechanisms of controlled and automatic processing are qualitatively different, the transition is a continuous process.

The connectionist simulation of automatic processing learns to perform visual search tasks. First, the model makes a few errors as it sets its performance criterion, then executes a slow serial search. As practice proceeds, it gradually acquires a fast parallel search. Connectionist autoassociative processing allows the network to generalize learning to similar patterns and provides an interpretation for why consistency is an important factor. The simulation of the model illustrates how a process can be both automatic and controlled and how the processes interact. It also has produced some novel predictions about cortical thalamic neural activity that are being examined physiologically.

EMPIRICAL IMPACT OF CONNECTIONISM

Although the theoretical impact of connectionism has been large, the empirical impact has been minimal and may remain limited. There is a very serious problem of the nonuniqueness of connectionist predictions. This problem is well illustrated by the modeling of the word superiority effect. McClelland and Rumelhart (1981) provided the archetype of a connectionist model that had three levels (a visual feature, a letter level, and a word level) to predict the word superiority effect. This model suggests that as an empiricist, one might try to perform experiments to examine the existence of each of these stages. However, Golden (1986) presented a model for the word superiority effect that had only a single level. In essence, he could predict the word superiority effect assuming only a visual feature level. The model did not even require a

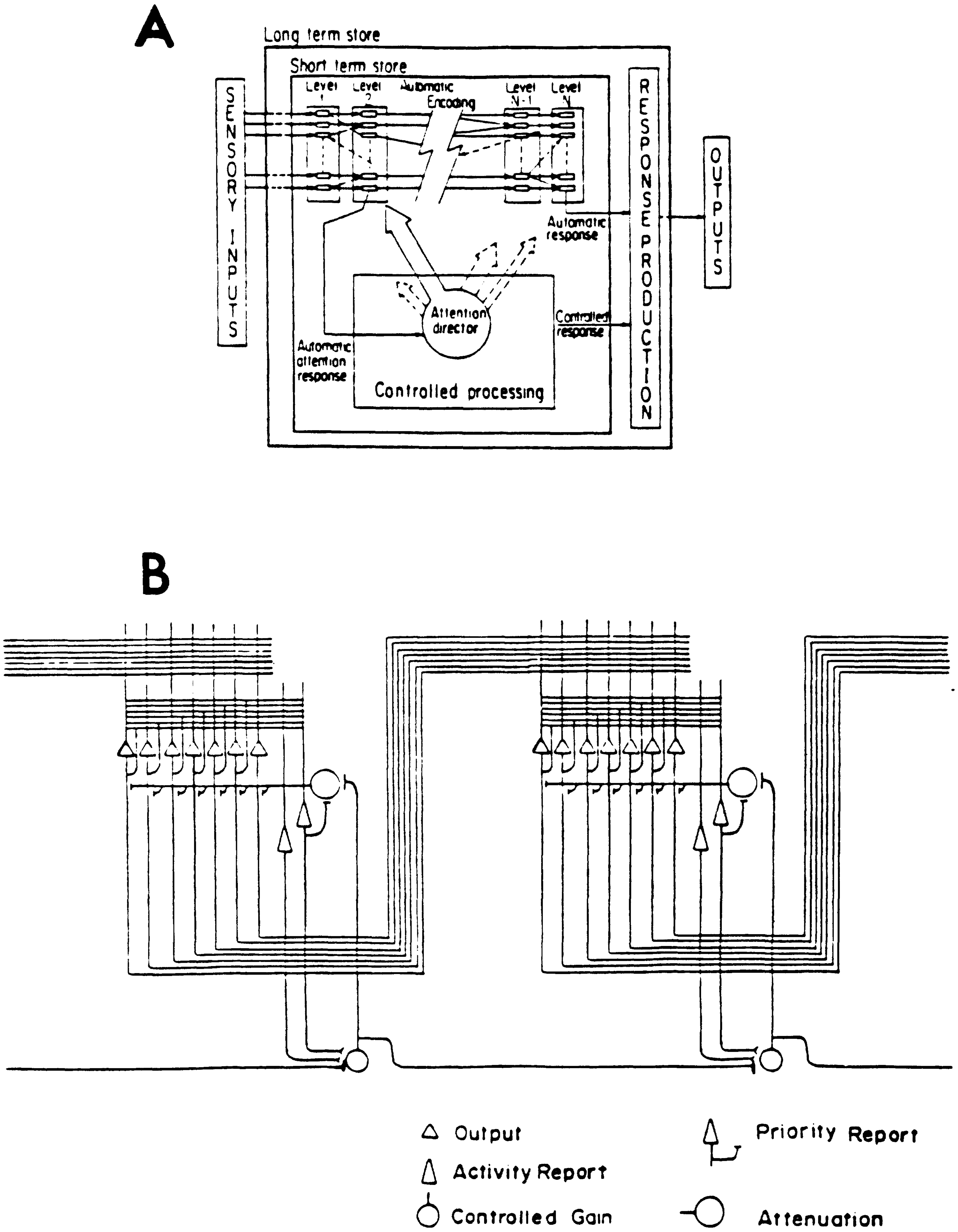

Figure 3A. Original Shiffrin and Schneider (1977, Figure 11) diagram for automatic and controlled processing. This represents a black box interpretation of attention and the interaction between controlled and automatic processing. The arrows between stages indicate the interactions between processes. For example, the solid arrow from a node in Level 2 to the attention system indicates an automatic attention response (AAR) causing attention to be shifted (large arrow). Figure 3B illustrates the Schneider and Mumme (1987) microstructure of a model that exhibits automatic and controlled processing as emergent features of the control architecture. The priority report cell in 3B conveys the same attention redirection as the AAR in 3A. However, in the current model the interaction between automatic processing (i.e., the local priority report to attention cell circuit) and controlled processing (i.e., the long distance activity/priority cell to controlled gain cell to attenuation cell circuit) is detailed. In the new model all the interactions are explicit. The model can learn to perform visual search tasks and produces human-like performance and learning curves. Figure 3A is from "Controlled and Automatic Human Information Processing-II. Perceptual Learning, Automatic Attending, and a General Theory," by R. M. Shiffrin and W. Schneider, 1977, *Psychological Review*, 84, p. 162. Copyright 1977 by the American Psychological Association.

visual letter level, much less a word level. The second connectionist model substantially countered the take-home message of the first connectionist model. The first model suggested that we should think in terms of top-down influences from the word level to the letter and visual feature level. The Golden model shows that we can have much the same effect, assuming there is nothing but a visual feature level of processing.

It is likely that within 5 years we will have a proliferation of connectionist models with very different architectures predicting the same empirical phenomena. Massaro (1986) presented a connectionist model that could predict a variety of effects in speech perception. Given the input and expected output, this system found connection weights that produced human-like data. Unfortunately, given slightly different output patterns, this system produces data that have never been observed in humans. It is critical to *remember that connectionist models use very powerful curve fitting procedures to map the input to the output.* Typically these models search in a several-thousand-parameter space of connections. These are powerful search techniques, and it is not surprising that they find solutions. This may be great for computer science, but causes a real problem for psychology. In general, psychologists seek to understand how humans perform processing. If 10 very different connectionist architectures can be built to model the same phenomenon, it is difficult to have much confidence in any one of the architectures. As connectionism matures, it will be critical to examine how it deals with this multiple-model problem. Mathematical psychology somewhat lost its enthusiasm because of its inability to resolve issues between models. In Norman's (1970) Models of Human Memory there were at least 12 different models for the recall curve. After the book was published, most of the contributors went on to perform different types of research, never coming to a consensus on the true underlying cause for the free recall effect.

When a connectionist model fails, there are many interpretations or outs for why it failed. Connectionist models are sensitive to the initial state, structure, number of elements, specific problems, learning sequence, learning rule, and coding patterns of the initial model. Given so many degrees of freedom and a very powerful learning rule, it is difficult to identify the limits of connectionist modeling. If the system fails to learn, there is always the possibility that given more units and more iterations, the system would have learned. Clear disconfirmation of a particular class of connectionist models is very hard to achieve.

WILL CONNECTIONISM FIZZLE?

It is important to note that perceptrons did fizzle. There was a great deal of early excitement, but after extensive analysis it was found that the learning systems were, in fact, far too limited. Connectionism is currently enjoying a very explosive growth, and it is hard to be rational during this period. To be viable, connectionism must deal with the problem of scaling well. The problem of scale is the bane of artificial intelligence. Many learning rules learn very well with small or toy problems but fail, due to a combinatoric explosion, with more complex problems. The scaling of connectionist models is not understood. Hinton indicates that they appear to scale by a factor of about N^3 to the number of connections. If it takes 10^4 learning trials to fill up a 100-connection network (as in NET-TALK), it would take 10^7 trials (or 14 man years of effort at 10 sec/trial) for a thousand-connection network. Cortical connection inputs can easily reach a million connections in a region. Connectionism must deal with procedures that allow problems to be decomposed so that the learning can occur in realistic time scales. Artificial intelligence started by generating great enthusiasm about general problem-solving methods. During this stage of artificial intelligence research, the mind was viewed as a tabula rasa. However, this approach quickly fell off a combinatoric cliff, making it untenable. Artificial intelligence started to solve real-world problems once it began trying to represent limited task domains via expert systems approaches. Some practitioners of connectionism feel that connectionism can solve the tabula rasa learning issue. My view is that eventually we will see some compromise between the position of restricted domain knowledge as an expert system and that of connectionism modeling to remove the brittleness of those systems. Norman (1986) comments that connectionism must deal with sequential processing, which is typical in human problem solving. To some extent, connectionist modeling can be viewed as modeling of events that typically occur in less than 1 sec. Much production system modeling (e.g., J. R. Anderson, 1983) looks at processing well above the 1-sec period. There is presently a great deal of interest in connectionism; however, one must be cautious that part of this enthusiasm may be coming from being tired of old concepts. Psychology dropped box models for semantic networks and production systems. It is now dropping those perhaps to embrace connectionism.

IS IT GOOD FOR THE FIELD?

Yes, but it may be another field. I generally think of psychology as being the study of human or animal systems. Connectionism studies learning systems that can be simulated in computers and may occur in animals. Human learning systems are a small sample of the possible learning systems that could exist. To make an analogy, think of the study of aerodynamics. To some extent, the study of aerodynamics began with the study of natural flight. Birds provided an existence proof of how an object could fly through the air under its own power. However, as the principles of aerodynamics began to be understood, researchers studied artificial man-made systems of flight. In cognitive science something similar may occur. Connectionist models may prove to be very effective learning systems that greatly advance the computation of learning. However, they may not perform those operations in a manner analogous to human learning.

CAUTIONS ON CONNECTIONISM

In a presidential address it is appropriate to comment about the status of the field. Although I view the connectionist movement with great enthusiasm, there are some factors that give me pause. Connectionism will produce some loss of the empirical tradition of psychology and perhaps promote an animosity toward other views. It is now acceptable to test learning concepts by running computer models as opposed to human subjects. This loosening of the paradigm is important and good for the field. However, I see developing signs of animosity between the modelers and the empirical researchers. If we are going to experience a paradigm shift, I hope that we can do it without the animosity that occurred as a result of Chomsky's linguistic theories. Chomsky's influential work caused many linguists to abandon the empirical study of linguistic processing in favor of the purely theoretical representation of that processing. The established connectionist modelers clearly have a strong regard for empirical data. I am, however, concerned by the younger generation of modelers, many of whom have only a passing interest in empirical data. I feel that if we wish to model human cognition, it is critical that we generate testable predictions so that we can limit the set of models that we search for.

HOW BIG A PARADIGM SHIFT?

I believe connectionist modeling does represent a significant paradigm shift in psychology. It is certainly beyond the level of a shift of the transition from box models to semantic nets in the early 1970s. Perhaps it is a shift approaching that of the shift from behaviorism to information processing in the late 1950s. It may be on a scale comparable to transformational grammar in linguistics. The current enthusiasm and exciting developments suggest that it may be the largest paradigm shift that most psychologists will see during their careers.

Connectionism is certainly changing the perspective that psychology has of human cognition. I end with a quote by Kuhn (1970, p. 121): "though the world does not change with a change in paradigm, the scientist afterward works in a different world."

REFERENCES

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**, 147-169.

Anderson, J. A. (1983). Cognitive and psychological computation with neural models. *IEEE Transactions on Systems, Man, & Cybernetics*, **13**, 799-815.

Anderson, J. A., & Mozer, M. C. (1981). Categorization and selective neurons. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 213-236). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Barto, A. G., & Anandan, P. (1985). Pattern recognizing stochastic learning automata. *IEEE Transactions on Systems, Man, & Cybernetics*, **15**, 360-375.

Elman, J., & Zipser, D. (1987). *Learning the hidden structure of speech* (Tech. Rep. No. ICS 8701). Institute for Cognitive Science, University of California, San Diego, CA.

Golden, R. M. (1986). A developmental neural model of visual word perception. *Cognitive Science*, **10**, 241-276.

Hammerstrom, D. (1986, August). *Neural computing: A new paradigm for LLSI computer architecture*. Paper given at the Attention and Brain Communication Workshop, Jackson, Wyoming.

Hinton, G. E. (1986). Learning distributed representations of concepts. *The eighth annual conference of the cognitive science society* (pp. 1-12). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Nowlan, S. J. (in press). *How learning can guide evolution* (Tech. Rep.). Carnegie-Mellon University, Pittsburgh, PA.

Hinton, G. E., & Plaut, O. C. (in press). *Using fast weights to deblur old memories and assimilate new ones* (Tech. Rep.). Carnegie-Mellon University, Pittsburgh, PA.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann Machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 282-317). Cambridge, MA: MIT Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Massaro, D. W. (1986, November). *Connectionist models of the mind*. Paper presented at Psychonomic Society meeting, New Orleans, LA.

McClelland, J. L. (1986). Resource requirements of standard and programmable nets. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition. Volume 1: Foundations* (pp. 460-487). Cambridge, MA: MIT Press.

McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 272-325). Cambridge, MA: MIT Press.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, **88**, 375-407.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.

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

Norman, D. A. (Ed.). (1970). *Models of human memory*. London: Academic Press.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2. Psychological and biological models* (pp. 531-546). Cambridge, MA: MIT Press.

Rosenberg, C. R., & Sejnowski, T. J. (1986). The spacing effect on NETtalk, a massively-parallel network. *The Eighth Annual Conference of the Cognitive Science Society* (pp. 72-89). Hillsdale, NJ: Erlbaum.

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

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Ex-*

plorations in the microstructure of cognition. Volume 1: Foundations (pp. 318-362). Cambridge, MA: MIT Press.

RUMELHART, D. E., & MCCLELLAND, J. L. (1986a). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216-271). Cambridge, MA: MIT Press.

RUMELHART, D. E., & MCCLELLAND, J. L. (Eds.). (1986b). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.

RUMELHART, D. E., SMOLENSKY, P., MCCLELLAND, J. L., & HINTON, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 7-57). Cambridge: MA: MIT Press.

SCHNEIDER, W., & DETWEILER, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21). New York: Academic Press.

SCHNEIDER, W., & MUMME, D. (1987). *Attention, automaticity and the capturing of knowledge: A two-level cognitive architecture.* Manuscript submitted for publication.

SEJNOWSKI, T. J. (1986). Open questions about computation in cerebral cortex. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2: Psychological and biological models* (pp. 372-389). Cambridge, MA: MIT Press.

SEJNOWSKI, T. J., & ROSENBERG, C. R. (1986). NETtalk: *A parallel network that learns to read aloud.* (Tech. Rep. No. JHU/EECS-86/01). The Johns Hopkins University Electrical Engineering and Computer Science, Baltimore, MD.

SHIFFRIN, R. M., & SCHNEIDER, W. (1977). Controlled and automatic human information processing. II: Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127-190.

WICKELGREN, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, **76**, 1-15.