March 2000
To appear in F. Boller and J. Grafman (Eds.), Handbook of Neuropsychology, 2nd Edition. New York: Elsiever.

Methodologies for the Computer Modeling of Human Cognitive Processes

David C. Plaut
Departments of Psychology and Computer Science
and the Center for the Neural Basis of Cognition
Carnegie Mellon University

Researchers interested in human cognitive processes have long used computer simulations to try to identify the principles of cognition. The strategy has been to build computational models that embody a set of principles and then the examine how well the models capture human performance in cognitive tasks. Computational modeling has both strengths and weaknesses in its usefulness in facilitating the development of cognitive theories. Perhaps its most important strength is that, in developing mechanistic theories of cognitive processes, it is critical to have a language or formalism for describing dynamic information processing in detail. Developing a working implementation of a model makes it possible both to verify the completeness and internal coherence of its underlying theoretical principles, and to generate detailed, quantitative predictions of the model when tested in novel circumstances. More generally, computational modeling can lead to new theoretical discoveries by providing a means of exploring the implications of a set of ideas or principles.

There are, however, a number of potential pitfalls in the application of computational modeling to cognitive processes that must be kept in mind. First, developing an explicit implementation always involves augmenting core theoretical claims with less central, ancillary assumptions, and it can be difficult to evaluate the extent to which the behavior of the model depends on the latter instead of the former. Moreover, models may be underconstrained and are applied to modeling data post hoc without a sufficient consideration of whether there are patterns of data that the model could *not* produce (see Roberts & Pashler, 2000). Finally, models are often put forward as theories in an of themselves, without sufficient analysis and explanation of *why* the model accounts for the data (see McCloskey, 1991). In general, computational modeling would seem to be most productive when it is carried out in the service of clarifying theoretical claims and when it it tightly integrated with corresponding empirical studies.

A number of formalisms have been applied to modeling cognitive processes, including production systems, discrimination nets, exemplar-based models, and connectionist models. The current chapter presents a brief summary of each of these approaches along with some illustrative examples. Particular emphasis is place on connectionist modeling as this is the framework that has been most widely applied to simulating neuropsychological phenomena.

Production Systems

Some of the earliest cognitive simulations, such as the General Problem Solver (Newell, Shaw, & Simon, 1958; Newell & Simon, 1972), took the form of production systems. In its general form, a production system consists of a set of condition-action rules, and a set of knowledge elements stored in *working memory*. When working memory elements satisfy the conditions of one or more rules, the rules "fire" and their actions are executed. These actions may involve adding or removing working memory elements, as well as gathering more input or generating behavioral output.

More recent cognitive models of this form have introduced particular elaborations of the basic production-system architecture. One such example is Soar (Newell, 1990). In Soar, productions are organized into *problem spaces* such that only the productions relevant to the current problem space can fire. Soar incorporates a principle of *impasse-driven subgoaling* whereby, if there is any ambiguity in what action to take next, an impasse is reached and a new problem space is created with the goal of resolving the impasse. Any impasse within this new problem space would spawn a further problem space, and so on, until the relevant actions are of a sufficiently fine grain that choice among them is straightforward. When the sequence of productions within a problem space succeeds in satisfying the space's goal, a new production or *chunk* is created that summarizes the conditions under which the particular action should be selected, thereby preventing the precipitating impasse in the future. In this way, chunking serves as a basic learning mechanism to progressive speed performance with practice. Soar has been applied to modeling a broad range of relatively high-level cognitive phenomena, including problem solving, analogy syntactic processes in language comprehension, and human-computer interaction (see Rosenbloom, Laird, & Newell, 1993).

Another influential framework for cognitive modeling based on production systems is ACT-R (Anderson, 1990; Anderson & Lebiere, 1998). Unlike Soar, ACT-R draws a sharp distinction between procedural and declarative knowledge. Procedural knowledge takes the form of sets of condition-action productions, much as in a standard production system. Learning new procedural knowledge involves *compilation* of new, more efficient productions based on the success patterns and dependencies of sequences of more primitive productions, similar to the chunking mechanism in Soar. Declarative knowledge in ACT-R, on the other hand, takes the form of a network of interconnected concept nodes; spreading activation among the nodes enables them to satisfy the conditions of productions (like active working memory elements in Soar). ACT-R has been applied primarily to modeling memory and categorization phenomena, although recent applications have focused on developing models of student learning in automated tutoring systems (Anderson, Corbett, Koedinger, & Pelletier, 1995). A closely related framework, 3CAPS (Just & Carpenter, 1992), applied mostly to language comprehension, explains individual differences in performance based on strict limitations in the amount of activation that can be propagated among working memory elements.

There have been relatively few attempts to apply production systems to modeling neuropsychological phenomena. One notable exception is Kimberg and Farah (1993), who accounted for the cognitive impairments of patients with frontal-lobe on motor sequencing, the Stroop task, the Wisconsin Cart Sorting Task, and a context memory task, in terms of limits in working memory capacity within a fairly traditional production system model.

Discrimination Nets

A discrimination net consists of a hierarchical tree of conditions or tests. Processing proceeds down the tree, with the result of each test directing processing along a particular sub-branch, until the appropriate categorization or action is determined by arriving at a leaf (end node) of the tree. Learning involves adding tests and reordering them to arrive at the appropriate leaf more efficiently. A well-known cognitive modeling framework based on discrimination nets is EPAM and its derivatives (Richman, Staszewski, & Simon, 1995), which has been applied successfully in accounting for a broad range of phenomena in perception, memory, and categorization. Discrimination-net algorithms have also been applied to language-related phenomena, including English inflectional morphology (Ling, 1994; Ling & Marinov, 1993). The approach has generally focused on understanding normal performance and has not as yet been applied to neuropsychological data.

Exemplar-Based Models

In an exemplar-based model, each categorized instance or example encountered by the system is explicitly stored, labeled by its category. A new input is categorized by determining its similarity to all stored exemplars, and assigning it the category of the most similar one, or the most common category among a number of similar ones. In addition to storing exemplars, many of the models can also learn to adjust weightings on input dimensions so as to allocate "attention" to the most relevant aspects of stimuli. Typical domains of application of exemplar models include memory (e.g., MINERVA, Hintzman, 1986), automatization (e.g., Logan, 1988), and category learning (e.g., ALCOVE, Kruschke, 1992, and the Generalized Context Model, Nosofsky, 1986). More recently, exemplar-based approaches are being adopted in the context of speech recognition (e.g., Goldinger, 1998) to account for long-term influences of surface properties of speech (e.g., talker gender) on recall accuracy. As with discrimination nets, exemplar-based models have focused largely on accounting for normal skilled performance rather than on the patterns of impaired performance that can arise from brain damage.

Connectionist Models

In connectionist models, cognitive processes take the form of cooperative and competitive interactions among large numbers of simple, neuron-like processing units. Typically, each unit has a real-valued activity level, roughly analgous to the firing rate of a neuron. Unit interactions are governed by weighted connections that encode the long-term knowledge of the system and are learned gradually through experience. The activity of some of the units encodes the input to the system; the resulting activity of other units encodes the system's response to that input. The patterns of activity of the remaining, so-called *hidden* units constitute learned, internal representations that mediate between inputs and outputs. Units and connections are not generally considered to be in one-to-one correspondence with actual neurons and synapses. Rather, connectionist systems attempt to capture the essential computational properties of the vast ensembles of real neuronal elements found in the brain through simulations of smaller networks of units. In this way, the approach is distinct from computational neuroscience (Sejnowski, Koch, & Churchland, 1988), which aims to model the detailed neurophysiology of relatively small groups of neurons. Although the connectionist approach uses physiological data to guide the search for underlying principles, it tends to focus more on overall system function or behavior, attempting to determine what principles of brain-style computation give rise to the cognitive phenomena observed in human behavior.

The simplest structure for a connectionist network is a *feedforward* architecture, in which information flows unidirectionally from input units to output units, typically via one or more layers of *hidden* units (so called because they are not visible to the environment). Although such networks can provide important insights into many cognitive domains, and are often reasonable approximations of more complex systems, they are severely limited in their ability to learn and process information over time and thus are relatively ill-suited for domains that involve complex temporal structure. A more appropriate type of network for such domains is a *recurrent* architecture that permits any pattern of interconnection among the units. In one type of recurrent network, termed an *attractor* network, units interact in such a way that, in response to a fixed input, the network as a whole gradually settles to a stable pattern of activity representing the network's interpretation of the input (including any associated response). Recurrent networks can also learn to process sequences of inputs and/or to produce sequences of outputs. For example, in a *simple recurrent* network (Elman, 1990), the internal representation generated by a given element of a sequence is made available as input to provide context for processing subsequent elements. Critically, the internal representations themselves adapt so as to provide and make effective use of this context information, enabling the system to learn to represent and retain relevant information across multiple time scales.

An issue of central relevance in the study of cognition is the nature of the underlying representation of information. Connectionist models divide roughly into two classes depending on the nature of the representations that are employed. In a *localist* representation, each unit corresponds to a distinct, familiar entity such as a letter, word, concept, or proposition (see Page, 2000). By contrast, in a *distributed* representation, such entities are encoded, not by individual units, but by alternative patterns of activity over the same group of units, so that each unit participates in representing many entities (see Hinton, McClelland, & Rumelhart, 1986). It is important to note that a representation is localist or distributed only with respect to a specific set of entities; the same representation can be localist with respect to some entities and distributed with respect to others. Both localist and distributed models are "connectionist" in the sense that the system's knowledge is encoded in terms of weights on connections between units (Feldman & Ballard, 1982).

Localist Models

A well-known example of a localist model is the Interactive Activation Model of letter and word perception (McClelland & Rumelhart, 1981), which consists of three layers of units: letter-feature units, letter units, and word units. Units in each layer receive inhibitory connections from mutually exclusive alternatives at the same letter (e.g., M vs. T in the first letter position); units receive facilitatory connections from units at other levels representing consistent information (e.g., M in the first letter position and the unit for MAKE at the word level). Interactive processing via these connections play a critical role in explaining a number of context effects in perception, such as the word superority effect (Reicher, 1969). Analogous models in the domains of spoken word perception (TRACE; McClelland & Elman, 1986) and production (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) have also been highly influential.

More generally, interactive localist models are often termed *constraint satisfaction* networks, as units represent alternative hypotheses and connection weights encode the constraints that govern what combinations of hypotheses constitute good solutions or interpretations. Constraint satisfaction networks have been applied effectively in modeling data on normal and impaired face recognition (Burton, Young, Bruce, Johnston, & Ellis, 1991), syntactic ambiguity resolution (Spivey & Tanenhaus, 1998), and analogical reasoning (Hummel & Holyoak, 1997; Thagard, 1989). An interesting subclass of models employs a second real value for each unit, analogous to the *phase* of periodic firing of a neuron. Synchrony among these values allows multiple stable coalitions of active units to be bound together without interfering with each other. Synchrony-binding networks have been applied to object recognition (Hummel & Biederman, 1992) and to inferential reasoning (Shastri & Ajjanagadde, 1993).

Because localist models stipulate the form and content of representations, they tend to de-emphasize the role of learning. Some localist models do, however, address learning phenomena. A notable example is Adaptive Resonance Theory (ART; Carpenter & Grossberg, 1987), which has been applied to a wide range of perceptual and cognitive domains, including some aspects of frontal lobe impairments (Levine & Prueitt, 1989). In ART, a *vigilance* parameter controls when the goodness-of-fit between the current stimuli and stored exemplars is sufficiently poor that a new exemplar must be created. A closely related framework, compound-cueing networks (Houghton & Tipper, 1996), has been applied to a number of sequential domains, including priming and immediate serial recall. In fact, many of the exemplar-based models (e.g., Kruschke, 1992) can be cast as localist connectionist models.

Another learning framework, falling somewhat between localist and distributed models, is Self-Organizing Maps, also known as Kohonen networks (see Kohonen, 1984). Internal units are organized topographically, usually in a two-dimensional sheet analogous to an area of cortex. The unit that responds most strongly to each input updates it weights to respond even more strongly to that input on subsequent presentations—a procedure known as competitive learning (Rumelhart & Zipser, 1985). In addition, the neighbors of the maximally responding unit also update their weights, biasing them to respond to similar stimuli. As a result, units become assigned to stimuli in such a way that the similarity structure among the inputs becomes layed out over the topographic organization of the units. Beyond being useful as a technique for dimensionality reduction and data visualization, Kohonen networks have been applied to a number of aspects of lexical, syntactic, and semantic processing (Miikkulainen, 1993, 1997; Schyns, 1991).

Distributed Models

Although localist models have been highly influential, many recent connectionist models rely on properties of distributed representations. Although such representations are more difficult to think about, they offer a rich and powerful basis for understanding learning, generalization, and the flexibility and productivity of cognition (Gelder, 1990). The key to the effectiveness of distributed representations is the use of patterns whose similarity relations correspond to the similarities in the roles the patterns play in cognition, given that similar patterns tend to have similar consequences in connectionist models (see Hinton et al., 1986). In very simple tasks, the similarities among the representations provided by the environment may be sufficient to guide behavior. However, in most cognitive domains, the functional relationships that must govern effective performance are often quite different from the similarities among surface forms. For example, the words CAT and CAP look and sound very similar but have entirely unrelated meanings. Consequently, the inputs to the system must be re-represented, perhaps via successive transformations across multiple intermediate layers of units, as new patterns of activity whose relative similarities abstract away from misleading surface similarity and, instead, capture the underlying structure of the domain. Distributed representations can also be used to implement more complex, relational knowledge structures like frames and scripts if units encode conjunctions of roles and properties of role-fillers—in fact, such representations emerge naturally when networks are trained on tasks in which entities enter into multiple types of relations (see Hinton, 1991).

Traditional approaches to understanding and modeling cognition typically make very strong and specific assumptions about the structure of these internal representations and the processes that manipulate them. The distributed connectionist approach, by contrast, places much greater emphasis on the ability of a system to *learn* effective internal representations. Learning in a connectionist network takes the form of modifying the values of weights on connections between units in response to feedback on the behavior of the network. A variety of specific learning procedures are employed in connectionist research; most that have been applied to cognitive domains, such as back-propagation (Rumelhart, Hinton, & Williams, 1986) take the form of error correction: Change each weight in a way that reduces the discrepancy between the correct response to each input and the one actually generated by the system. In this process, internal representations over hidden units are learned by calculating how to change each unit's activation to reduce error and then modifying its incoming weights accordingly. Although it is unlikely that the brain implements back-propagation in any direct sense (Crick, 1989), there are more biologically plausible procedures that are computationally equivalent (see, e.g., O'Reilly, 1996).

The emphasis on learning within the connectionist approach has fundamental implications for the nature of the explanations offered for cognitive behavior. Instead of attempting to stipulate the specific form and content of the knowledge required for performance in a domain, the approach instead stipulates the *tasks* the system must perform, including the nature of the relevant information in the environment, but then leaves it up to learning to develop the necessary internal representations and processes (McClelland, St.John, & Taraban, 1989). In some contexts, the resulting solution may bear a close relationship to more traditional mechanisms, but it is more often the case that learning develops representations and processes which are radically different from those proposed by traditional theories, and which generate novel hypotheses and testable predictions concerning human cognitive behavior. Distributed connectionist models have been applied to the full range of perceptual, cognitive, and motor domains (see McClelland, Rumelhart, & PDP ResearchGroup, 1986; Quinlan, 1991; McLeod, Plunkett, & Rolls, 1998).

In an early application of error-correcting learning, Rumelhart and McClelland (1986) showed that a single network could learn to generate the past-tense forms of both regular and irregular English verbs from their stems, thereby obviating the need for dual rule-based and exception mechanisms (Pinker, 1984, 1991). Although aspects of the approach were strongly criticized (Pinker & Prince, 1988), many of the specific limitations of the model have been addressed in subsequent simulation work (see, e.g., MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993, 1996). More recently, Joanisse and Seidenberg (1999) developed a connectionist model of past-tense formation in which representations of input phonology, output phonology, and semantics all interact in performing lexical tasks. When the network was tested on past-tense generation, semantic damage impaired performance more for irregular verbs than for regular verbs or novel verbs, whereas phonological damage impaired performance more for novel verbs than for regular or irregular verbs. These findings correspond to the dissociations observed by Ullman et al. (1997) with Alzheimer's disease and Parkinson's disease, respectively, and interpreted by them as implicating separate rule-based and lexical mechanisms.

A similar line of progress has taken place in the domain of English word reading. An early connectionist model (Seidenberg & McClelland, 1989) provided a good account of word reading but was poor at pronouncing word-like pseudowords (e.g., MAVE; Besner, Twilley, McCann, & Seergobin, 1990). A more recent series of simulations (Plaut, McClelland, Seidenberg, & Patterson, 1996) showed that the limitations of this preliminary model stemmed not from any general failing of connectionist networks, but from the model's use of poorly structured orthographic and phonological representations. When representations were used that condense the regularities between orthography and phonology by incorporating graphotactic and phonotactic constraints, networks were able to learn to pronounce both regular and exception words, and yet also pronounce pseudowords as well as skilled readers. When the models are damaged in various ways, they exhibit the major forms of acquired dyslexia, including *deep* dyslexia in which patients make semantic errors in reading aloud (e.g., misreading YACHT as "boat"; Hinton & Shallice, 1991; Plaut & Shallice, 1993), and *surface* dyslexia in which patients produce regularization errors to exception words (e.g., misreading YACHT as "yatched"; Plaut, 1997; Plaut et al., 1996). Moreover, retraining the damaged models yields patterns of recovery and generalization that are qualitatively similar to those found in cognitive rehabilitation studies (Plaut, 1996). A fully recurrent network operating according to very similar principles, but with an additional attentional mechanism, has been used to account for the interaction of both perceptual and lexical/semantic factors in the reading errors of neglect dyslexic patients (Mozer & Behrmann, 1990).

Although fully recurrent networks are capable of learning to exhibit complex temporal behavior, for reasons of efficiency it is more common to apply simple recurrent networks in temporal domains. For example, Elman (1991) demonstrated that a simple recurrent network could learn the structure of an English-like grammar, involving number agreement and variable verb argument structure across multiple levels of embedding, by repeatedly attempting to predict the next word in processing sentences (see also Rohde & Plaut, 1999). St.John and McClelland (1990) also showed, for a somewhat simpler corpus, how such networks can learn to develop a representation of sentence meaning by attempting to answer queries thematic role assignments throughout the course of processing a sentence. Cleeremans and McClelland (1991) demonstrated that simple recurrent networks could account for a number of phenomena related to implicit learning of sequential structure generated by artificial grammars.

One of the main attractions of distributed connectionist models such as these is their ability to discover the structure implicit in ensembles of events and experiences (see also McClelland & Rumelhart, 1985). Accomplishing this, however, requires making only very small changes in response to each input so that the resulting weight values reflect the long-term experience of the system. Attempts to teach such networks the idiosyncratic properties of specific events one after the other do not generally succeed since the changes made in learning each new case produce "catastrophic interference" with what was previously stored in the weights (McCloskey & Cohen, 1989). McClelland, McNaughton, and O'Reilly (1995) observed, however, that catastrophic interference does not occur if continued training of old knowledge is interleaved with the training of new knowledge. They proposed that the brain employs two complementary learning systems: a cortical system for gradual learning using highly overlapping distributed representations, and a subcortical, hippocampal-based system for rapid learning using much sparser, less-overlapping representations. On their account, stored instances in the hippocampus provide the training input for past experience that must be interleaved with ongoing experience to prevent interference in cortex. The argument was that learning in cortex and in distributed networks are similarly constrained, so that the strengths and limitations of structure-sensitive learning in networks explained *why* the brain employs two complementary learning systems in hippocampus and neocortex.

Summary

A broad range of computational formalisms have been employed in the service of developing simulations of human cognitive processes. Some of these, such as production systems and discrimination nets, are based on the application of symbolic rules, whereas others, such as distributed connectionist networks, are based on putative principles of neural computation. While the former have provided important insights into high-level cognitive processes such as problem solving and reasoning, the latter have proven more useful in understanding how disorders of brain function can lead to disorders of cognition. Although computational modeling offers important tools for theory development, it is most useful when computational and empirical studies are carred out in an integrated, mutually informing manner.

References

Anderson, JR: *The adaptive character of thought.* Hillsdale, NJ: Erlbaum, 1990.

Anderson, JR, Corbett, AT, Koedinger, KR, Pelletier, R: Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167-207, 1995.

Anderson, JR, Lebiere, C: *The atomic components of thought.* Mahwah, NJ: Erlbaum, 1998.

Besner, D, Twilley, L, McCann, RS, Seergobin, K: On the connection between connectionism and data: Are a few words necessary? *Psychological Review, 97*, 432-446, 1990.

Burton, MA, Young, AW, Bruce, V, Johnston, RA, Ellis, AW: Understanding covert recognition. *Cognition, 39*, 129-166, 1991.

Carpenter, GA, Grossberg, S: A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37*, 54-115, 1987.

Cleeremans, A, McClelland, JL: Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235-253, 1991.

Crick, F HC: The recent excitement about neural networks. *Nature, 337*, 129-132, 1989.

Dell, GS: A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321, 1986.

Dell, GS, Schwartz, MF, Martin, N, Saffran, EM, Gagnon, DA: Lexical access in normal and aphasic speakers. *Psychological Review, 104*, 801-838, 1997.

Elman, JL: Finding structure in time. *Cognitive Science, 14*, 179-211, 1990.

Elman, JL: Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195-225, 1991.

Feldman, JA, Ballard, DH: Connectionist models and their properties. *Cognitive Science, 6*, 205-254, 1982.

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

Goldinger, SD: Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279, 1998.

Hinton, GE (Ed.).: *Connectionist symbol processing.* Cambridge, MA: MIT Press, 1991.

Hinton, GE, McClelland, JL, Rumelhart, DE: Distributed representations. In DE Rumelhart, JL McClelland, PDP ResearchGroup (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (p. 77-109). Cambridge, MA: MIT Press, 1986.

Hinton, GE, Shallice, T: Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74-95, 1991.

Hintzman, DL: "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411-428, 1986.

Houghton, G, Tipper, SP: Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action. *Brain and Cognition*, *30*, 20-43, 1996.

Hummel, JE, Biederman, I: Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480-517, 1992.

Hummel, JE, Holyoak, KJ: Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427-466, 1997.

Joanisse, MF, Seidenberg, MS: Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Science, USA*, *96*, 7592-7597, 1999.

Just, MA, Carpenter, PA: A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149, 1992.

Kimberg, DY, Farah, MJ: A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*, 411-428, 1993.

Kohonen, T: *Self-organization and associative memory.* New York: Springer-Verlag, 1984.

Kruschke, JK: ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44, 1992.

Levine, DS, Prueitt, PS: Modeling some effects of frontal lobe damage—Novelty and perseveration. *Neural Networks*, *2*, 103-116, 1989.

Ling, CX: Learning the past tense of English verbs: The symbolic pattern associator vs. connectionist models. *Journal of Artificial Intelligence Research*, *1*, 209-229, 1994.

Ling, CX, Marinov, M: Answering the connectionist challenge: A symbolic model of learning the past tense of English verbs. *Cognition*, *49*, 235-290, 1993.

Logan, GD: Toward and instance theory of automatization. *Psychological Review*, *95*, 492-527, 1988.

MacWhinney, B, Leinbach, J: Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*, 121-153, 1991.

McClelland, JL, Elman, JL: The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86, 1986.

McClelland, JL, McNaughton, BL, O'Reilly, RC: Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419-457, 1995.

McClelland, JL, Rumelhart, DE: An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407, 1981.

McClelland, JL, Rumelhart, DE: Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159-188, 1985.

McClelland, JL, Rumelhart, DE, PDP ResearchGroup the (Eds.).: *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press, 1986.

McClelland, JL, St.John, M, Taraban, R: Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*, 287-335, 1989.

McCloskey, M: Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, *2*, 387-395, 1991.

McCloskey, M, Cohen, NJ: Catastrophic interference in connectionist networks: The sequential learning problem. In GH Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, p. 109-165). New York: Academic Press, 1989.

McLeod, P, Plunkett, K, Rolls, ET: *Introduction to connectionist modelling of cognitive processes.* Oxford, UK: Oxford University Press, 1998.

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

Miikkulainen, R: Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, *59*, 334-366, 1997.

Mozer, MC, Behrmann, M: On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*, 96-123, 1990.

Newell, A: *Unified theories of cognition.* Cambridge, MA: MIT Press, 1990.

Newell, A, Shaw, JC, Simon, HA: Elements of a theory of human problem solving. *Psychological Review*, *65*, 151-166, 1958.

Newell, A, Simon, HA: *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall, 1972.

Nosofsky, RM: Attention, similiarity and the identification-categorization relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *115*, 39-57, 1986.

O'Reilly, RC: Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*, 895-938, 1996.

Page, M: Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, 2000.

Pinker, S: *Language learnability and language development.* Cambridge, MA: Harvard University Press, 1984.

Pinker, S: Rules of language. *Science*, *253*, 530-535, 1991.

Pinker, S, Prince, A: On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73-193, 1988.

Plaut, DC: Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, *52*, 25-82, 1996.

Plaut, DC: Structure and function in the lexical system: Insights from distributed models of naming and lexical decision. *Language and Cognitive Processes*, *12*, 767-808, 1997.

Plaut, DC, McClelland, JL, Seidenberg, MS, Patterson, K: Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115, 1996.

Plaut, DC, Shallice, T: Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377-500, 1993.

Plunkett, K, Marchman, VA: From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21-69, 1993.

Plunkett, K, Marchman, VA: Learning from a connectionist model of the acquisition of the English past tense. *Cognition*, *61*, 299-308, 1996.

Quinlan, P: *Connectionism and psychology: A psychological perspective on new connectionist research.* Chicago: University of Chicago Press, 1991.

Reicher, GM: Perceptual recognition as a functionof meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274-280, 1969.

Richman, HB, Staszewski, JJ, Simon, HA: Simulation of expert memory using EPAM IV. *Psychological Review*, *102*, 305-330, 1995.

Roberts, S, Pashler, H: How persuasive is a good fit? A comment on theory testing in psychology. *Psychological Review*, 2000.

Rohde, D LT, Plaut, DC: Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67-109, 1999.

Rosenbloom, PS, Laird, JE, Newell, A (Eds.).: *The Soar papers: Research on integrated intelligence, volumes 1 and 2.* Cambridge, MA: MIT Press, 1993.

Rumelhart, DE, Hinton, GE, Williams, RJ: Learning representations by back-propagating errors. *Nature*, *323*, 533-536, 1986.

Rumelhart, DE, McClelland, JL: On learning the past tenses of English verbs. In JL McClelland, DE Rumelhart, PDP ResearchGroup (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (p. 216-271). Cambridge, MA: MIT Press, 1986.

Rumelhart, DE, Zipser, D: Feature discovery by competitive learning. *Cognitive Science*, *9*, 75-112, 1985.

Schyns, PG: A modular neural network model of concept acquisition. *Cognitive Science*, *15*, 461-508, 1991.

Seidenberg, MS, McClelland, JL: A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568, 1989.

Sejnowski, TJ, Koch, C, Churchland, PS: Computational neuroscience. *Science*, *241*, 1299-1306, 1988.

Shastri, L, Ajjanagadde, V: From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, *16*, 417-494, 1993.

Spivey, MJ, Tanenhaus, MK: Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 1521-1543, 1998.

St.John, MF, McClelland, JL: Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217-257, 1990.

Thagard, P: Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435-502, 1989.

Ullman, MT, Corkin, S, Coppola, M, Hicock, G, Growdon, JH, Koroshetz, WJ, Pinker, S: A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*, 266-276, 1997.