Computational Approaches to Cognition: Top-Down Approaches

James L. McClelland and **David C. Plaut**
Department of Psychology
Carnegie Mellon University

Current Opinion in Neurobiology, *3*, 209–216.
Special Issue on Cognitive Neuroscience

Summary

Computational models offer tools for exploring the nature of human cognitive processes. In this article we focus on connectionist models, since these models are suggesting new ways of thinking about the basic nature of cognition and its implementation in the brain. The models support novel explanations of important aspects of perception, memory, language, thought, and cognitive development. They also provide tools for linking cognitive processes with the underlying physiological mechanisms, and for understanding how disorders of brain function lead to disorders of cognition.

Introduction

Since the pioneering work of Newell and Simon in the late 1950's, researchers interested in human cognitive processes have used computer simulations to try to identify the principles of cognition. The strategy has been to build computational models that embody putative principles and then to examine how well such models capture human performance in cognitive tasks. Until the 1980's, this effort was undertaken within the context of the 'computer metaphor' of mind. Researchers thought of the human mind as though it were a conventional digital computer and built computational models based on this conceptualization. However, in the late 70's and early 80's this began to change [1,2,3]. By the mid-1980's research exploring principles that appear to be consistent with the style of computation employed by the brain was in full swing [4,5]. Recent developments in this new style of computational modeling, often called connectionist or neural network modeling, or the parallel distributed processing approach, will be the focus of this article.

In connectionist models, processing is generally assumed to emerge from the interactions of large numbers of simple neuron-like computational elements generally called units, and communication among the units is assumed to be based on plastic excitatory and inhibitory connections among the units. While each unit exhibits non-linear spatial and temporal summation, units and connections should not generally be taken as standing in one-to-one correspondence to actual neurons and synapses. Rather, the approach attempts 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.

An issue of central relevance is the nature of the representations used in cognitive processes. Some connectionist models use localist representations, in which individual units stand for familiar entities such as letters, words, concepts, and propositions. Others use distributed representations in which, in general, such entities are represented by a pattern of activity rather than by the

activity of a single unit. The key to the use of distributed representations is the use of patterns whose similarity relations capture similarities in the roles the patterns play in cognition, since, in connectionist models, similar patterns have similar consequences (see [6] for a full discussion).

The research within the connectionist framework may be considered 'top-down' since it tends to focus on overall system function or behavior, and asks: What principles of brain-style computation give rise to the behavioral phenomena that we observe in human cognition. Because the approach exploits principles thought to be consistent with the style of information processing in the brain, it has also proven useful for developing ideas about the neural implementation of human cognition and for interpreting data arising from biological disorders of cognition. The approach might best be called interactive, rather than top-down, however, in that it often uses both behavioral and physiological data to guide the search for underlying principles. The approach has also been called an abstract or abstractive approach [7], since it seeks to identify the crucial principles of processing and does this by examining how far one may go in accounting for phenomena of interest while still abstracting from some neurophysiological details.

Computational Advances

In attempting to elucidate some of the general computational principles that characterize human cognitive function, various investigators have focused on the adaptive character both of cognition and of the underlying biological substrate (primarily synaptic plasticity). Previous developments [8,9] centered on the discovery of powerful training procedures that enable connectionist networks to develop useful internal representations via some form of error-correcting learning. An important recent development in this arena has been the discovery of how to do error- correcting learning when the error information is given in terms of distal stimulus coordinates [10••]. This is a major advance, since it takes away the need for the 'teacher' to guide the actual output variables directly. Instead it may simply provide information about how well the output served to produce an intended effect, such as how well the motor commands for a particular jump- shot succeeded in putting a basketball through a hoop.

There has also been a focus on using criteria other than error as a basis for connection adjustment. For example, Linsker [11] has proposed that the goal of connection adjustment may be to maximize the information that one population of units retains about the pattern of activation on another population. Others have proposed minimizing not only error but also some measure of the complexity of the network, such as the magnitude and/or number of connection weights [12•] or the number of long-range connections [13]. This work is of considerable interest because the use of these kinds of minimization functions gives rise to superior generalization performance after learning and promotes the development of patterns of connectivity characteristic of neural developmental processes [14,15•]. Jacobs and colleagues [16•,17•] have also demonstrated improved learning and generalization within a modular network architecture in which separate expert networks become specialized for particular subtasks.

Other work has attempted to characterize the general principles of information processing beyond its adaptive character. The approach one of us has taken has been to try to enumerate the principles that apply broadly across domains of cognitive function—principles that must be embodied in any specific computational model that attempts to account for performance in a particular domain. At present some of the candidate principles under active exploration are the principles of cascaded, stochastic and interactive propagation of information [18]. Of these, the principle of interactive processing (bi-directional propagation of information) has been the subject of the most intense scrutiny. Some researchers have argued that interactivity is inconsistent with robust characteristics of human performance in the very tasks that in the past were taken to implicate

interactive processing [19]. However, [20•] has established that in fact stochastic interactive processing is fully consistent with the problematic data. The problem with the old models was that they were deterministic, not that they were interactive.

The following sections consider applications of connectionist models to specific content areas in human cognition. In many areas, we will see that connectionist models suggest alternatives to conventional conceptions of the nature of the underlying cognitive processes and their neural implementation.

Language

Human language processing and acquisition has been a focus of connectionist research for many years. Controversy continues to swirl around the claims that have arisen from earlier connectionist models. A central claim arising from a relatively early distributed connectionist model [21] has been that language processing can dispense with explicit rules and lists of exceptions and special cases, and rely instead on gradual adaptation of connection weights to capture both the general regularities and the idiosyncratic properties of particular items. Many connectionist models have demonstrated considerable success in capturing aspects of morphology [22,23,24,25], and one recent paper has shown how one case of a gradual process of language change can be accounted for by the successive regularizing effects of several generations of neural networks [26•]. Furthermore, Elman [27••] has extended the approach to the syntactic level by showing how a simple recurrent network can learn the structure of an English-like grammar involving number agreement and verb argument structure, across multiple levels of embedding. Some researchers still maintain there are reasons to adhere to the classical rules-plus-exceptions story [28]. Others have attempted to integrate aspects of both connectionist and symbolic approaches in efforts to construct fast and efficient systems for acquiring and using lexical representations for language processing [29,30].

Some of the issues about the basic nature of information processing that were discussed above have arisen in specific areas in psycholinguistics. In particular, there has been active exploration of whether processing occurs in discrete, feedforward stages, or whether it is cascaded and interactive (cf. [18]). Levelt and colleagues [31] have argued strongly against the latter view on the basis of the lack of mediated semantic-to-phonological priming in picture naming (e.g., CAT primes DOG but not a phonological neighbor of DOG, LOG). Dell and O'Seaghdha [32•] have recently shown that the lack of mediated priming, along with significant mixed semantic-phonological priming (e.g., CAT primes RAT more than DOG), can be accommodated within an interactive connectionist system so long as words activate semantic and phonological neighbors only to a relatively small degree. However, Levelt and colleagues also found no evidence of feedback from phonology to semantics at later stages of processing, leading Dell and O'Seaghdha to suggest that the language production system is globally modular but locally interactive.

The model on which Dell and O'Seaghdha's work is based [33] follows traditional accounts of speech production in that phonological structure is explicitly reflected in the structure of the model (e.g., units for syllables and rimes). Dell, Juliano, and Govindjee [34•] demonstrate that distributed networks that develop internal representations can learn the same phonological structure – as evidenced by patterns of speech errors – based only on exposure to the statistical properties of pronunciations.

Cognitive Development

A central issue in cognitive development is the relative contribution of innate structure, maturational change, and learning to the pattern of development of cognitive processes. Several in-

vestigators have recently begun to explore the extent to which the outcome and the time course of development may be accounted for in terms of experience-based connection adaptation processes, shaped either by initial network architecture and parameters or by gradual maturational or learning-based changes (see [35•] for a review; other general discussions of the implications of connectionist models for issues in development may be found in [36,37,38]). Here we focus on two specific modeling studies that suggest how the effects of maturation and learning in networks may account for data that might otherwise seem to implicate highly-specific innate structure.

Elman tested the hypothesis that immaturity in the learning mechanism, in the form of reduced short-term memory, might actually help rather than hinder language acquisition (also see [39]). As previously mentioned, Elman [27••] trained a simple recurrent network to predict the syntactic structure of English-like sentences. In further studies [40••,41••], he found that the network succeeded at learning the task if its short-term memory was initially limited but gradually improved, whereas it failed if given fully mature short-term memory at the outset. Elman argued that maturational change compensates for inherent limitations in the learning ability of networks. In essence, the immature system is sensitive to only the most salient sources of variation in the task, allowing it to ignore more subtle distinctions until after it has mastered the basics.

However, work by Marchman [42••] suggests that some 'critical period' effects can arise solely from the nature of learning in networks without specific maturational change. She studied the effects of lesioning a network at various points during the acquisition of the English past tense. When learning only regular forms (e.g., TALK $\Rightarrow$ TALKED), long-term performance was relatively unaffected by lesions, although recovery took longer after lesions that were more severe or occurred later in training. As the network gains experience with the task, its knowledge becomes more 'entrenched,' making it less able to adapt to the effects of damage. In studies involving training a network on both regular and irregular forms (e.g., GO $\Rightarrow$ WENT), the regular forms were slower to learn and more disrupted by lesions, whereas the irregular forms, owing to their generally higher frequency, were learned quickly and were relatively insensitive to damage. Thus, a dissociation of performance on regular vs. irregular forms can arise from a generalized impairment, and so does not provide evidence for separate 'rule-based' and 'associative' mechanisms in language processing (cf. [28]).

Considerable research has identified constraints on how children form concepts and associate them with names (see, e.g., [43]). Schyns [44•] has shown how many of these constraints are satisfied naturally within a modular connectionist architecture in which bottom-up, unsupervised learning of concepts interacts with top- down, supervised association of these concepts with arbitrary labels. Of particular interest is the demonstration that top-down influences can elaborate those lower-level distinctions that are required for naming.

Learning and Memory

There have been a number of applications of connectionist models to topics in human memory. Ironically, though, connectionist models that are excellent for capturing cognitive development by gradually discovering the structure implicit in ensembles of events and experiences (e.g., [27••,45]) are not so good at learning the specific contents of events and experiences that must be learned in succession [46]. Networks that gradually discover structure do so by utilizing overlapping patterns of activation for concepts that the cognitive system has discovered have similar meanings or implications. Such networks discover these representations gradually by making only tiny adaptive changes in response to individual events and experiences. Attempts to teach such networks the specific idiosyncratic properties of specific events one after the other do not succeed, since the changes made in learning each new case produce catastrophic interference with what was previously stored

in the connection weights. For this reason, successful models of human learning and memory tend to use a coding that to maximizes the differentiation of similar events. For example, one version of Kruschke's ALCOVE [47] essentially assigns a single connectionist unit to represent each entire event or memory. This minimizes interference since the connection weights involved in storing different memories are completely non-overlapping.

One might seek a single model that combines the ability to gradually discover how to represent events and experiences using distributed representations with the ability to capture the idiosyncratic properties of specific events and experiences. However, the literature on human amnesia suggests that this may be the wrong approach. Research over the last 30 years has demonstrated that bilateral removal of or damage to the hippocampus and related structures in the temporal lobes produces a profound deficit in the ability to learn new episodic information and to recall information recently acquired, but appears to leave intact the ability to learn skills or to exhibit subtle priming effects of specific events and experiences. Taken together with the successes and failures of distributed connectionist models reviewed above, this dissociation leads to the suggestion that perhaps the hippocampus contains a memory system similar to ALCOVE for episodic learning, leaving the cortex free for the gradual discovery of useful distributed representations through interleaved learning [48•]. This approach has the virtue of explaining the curious phenomenon of temporally graded retrograde amnesia—the fact that hippocampal amnesics exhibit a selective deficit for recent, but not remote, episodic information. On the two-system view, temporally graded retrograde amnesia reflects the gradual incorporation of material initially learned by the hippocampus into the cortical memory system. For alternative computational models of hippocampal function, including detailed application to the animal classical conditioning literature, see [49•,50].

Visual Cognition

People can easily recognize objects despite changes in viewpoint, and yet understanding the mechanism by which this is done remains a challenging research problem. Olshausen, Anderson, and Van Essen [51•] address this problem in the context of developing a model of visual attention that makes specific contact with neuroanatomical and neurophysiological data. In their model, attention implements a 'window' on the visual field, selecting a particular region of lower-level cortical activity to be routed onto the same higher cortical area, retaining the relative spatial arrangement of image features. Olshausen and colleagues show how control signals can dynamically adjust the position and size of the window so that an object anywhere in the visual field activates the same higher-level units, allowing it to be recognized by a form of template matching.

Hummel and Biederman [52••] take a somewhat different approach to the problem of viewpoint-invariant object recognition. Rather than retain precise relative spatial information, their network derives only categorical relationships (e.g., ABOVE, BESIDE) among volumetric primitives called 'geons' [53] (see [54] for simulations supporting a distinction between categorical and coordinate spatial representations). Specialized connections in the lower levels of the network parse the image into geons by synchronizing the oscillatory outputs of units representing local image features. At intermediate levels, the same phase locking mechanism binds units representing relationships among geons to the attributes of the geons themselves. The bound attributes and relationships constitute a viewpoint-invariant structural description that is recognized by one of a set of object-specific units at the highest level of the network.

The extent that these oscillations are available, and the extent that they are relied upon in perception and cognition remains to be established. The use of temporal synchrony for dynamic feature binding is based on recent controversial findings of stimulus- specific synchronized oscillations of cortical activity [55]. Mozer, Zemel, Behrmann, and Williams [56] present a proposal for

how such synchronization could be learned.

Higher Cognition

People bring a vast amount of knowledge to bear in making implicit inferences during language understanding and commonsense reasoning. Two connectionist approaches to explaining this ability have recently been elaborated in considerable detail: the first within a localist framework; the second within a more distributed framework.

Efficient implicit reasoning requires representing and manipulating dynamic bindings between concepts and their roles in particular contexts. For example, the meaning of the sentence 'JOHN GAVE MARY THE BOOK' can be broken down into the bindings JOHN/ giver, MARY/recipient, and BOOK/object. Shastri and Ajjanagadde [57•] represent concepts and roles as individual units and—like Hummel and Biederman [52••]—use synchronous firing to reflect temporary bindings among them. Long-term knowledge (e.g., the act of giving involves a change of ownership) is represented and applied by subnetworks that detect synchrony among some units and cause it among others (e.g., MARY/owner/BOOK). In this way, reasoning involves the transient, systematic propagation of synchronized activity among concepts and roles.

In contrast, Smolensky, Legendre, and Miyata [58•] represent concepts and roles as distributed patterns of activity over large groups of units, and use a tensor product formalism—a generalization of the outer product of matrix algebra—to represent concept/role bindings. At the lower level, mental processes consist of massively parallel spreading activation; at a higher level, the same processes constitute a form of symbol manipulation in which entire structures are manipulated in parallel. The tensor product formalism enables a unified characterization of these different levels of analysis in terms of the optimization of well-formedness, or 'Harmony'. The approach is useful in accounting for some complex interactions among semantic and syntactic factors in language understanding, as well as in providing a novel account of the systematicity and productivity of language and thought (contra [59]). However, more adaptive representational schemes than those envisioned either in [58•] or [57•] may be necessary to capture human representational capacities in a way that is both efficient and sensitive to domain-relevant structure.

Cognitive Neuropsychology

Connectionist models are leading to new ways of understanding the possible implications of cognitive disorders for our understanding of normal cognitive function. In the past, investigators have often proceeded by stipulating the existence of a specialized module in the cognitive system whenever a brain-damaged patient exhibits a selective deficit in some specific aspect of cognitive function. For example, patients with 'hemispatial neglect' are abnormally slow to shift their attention from a pre-cued ipsilesional location to a contralesional stimulus. This has been interpreted in terms of damage to a specific 'disengage' module [60]. Connectionist models are leading to re-evaluations of many such inferences. In this case, for example, Cohen, Romero, Servan-Schreiber, and Farah [61•] reproduced the deficit in shifting attention by unilaterally damaging a network in which attention is allocated based on competitive interactions among units representing different spatial locations, demonstrating that no special disengage module in needed.

In a similar way, connectionist models are contributing to the question of whether knowledge is organized in the cognitive system by modality or by category. Apparent evidence for category specificity comes from the fact that some brain-injured patients are selectively impaired in recognizing and recalling information about living vs. nonliving things. Warrington and Shallice [62] suggested that these deficits could be accounted for if semantics were instead organized by modal-

ity, under the assumption that visual semantics is impaired and living things rely more heavily on visual than on functional semantics. In the absence of a computational model, this account was rejected in favor of the idea that knowledge is organized by category when it was noted that patients had difficulty both with functional as well as visual aspects of living things. But Farah and McClelland [63•] revived the original Warrington and Shallice hypothesis by showing that they could account for the entire pattern of deficits in an interactive and distributed model in which visual and functional semantics interact with each other as well as with visual and verbal input. The simulation accounted for the patients' paradoxical impairment in recalling functional information about living things because functional semantics relies on interactions with visual semantics to settle to the correct pattern.

As another example of the influence of connectionist models on interpretation of neuropsychological deficits, patients with 'prosopagnosia' fail to name familiar faces and have no conscious recollection of them, and yet often show evidence of recognition on more covert tasks, such as priming or name relearning, suggesting to some separate mechanisms of overt and covert recognition. But simulations by Burton, Young, Bruce, Johnston, and Ellis [64•] within a localist framework, and by Farah, O'Reilly, and Vecera [65•] within a distributed framework, demonstrate that this dissociation between overt and covert face recognition can arise in a single system from residual information after partial damage, eliminating any need to postulate separate mechanisms for recognition and for conscious awareness.

Principles of connectionist representations and processes can provide insight, not only into the overall organization of the cognitive system, but also into the specific patterns of errors that can arise from brain damage. In one important example of this approach, it has been established that robust characteristics of attractor networks (that is, interactive networks that use distributed representations and settle to sable attractors) provide parsimonious accounts of the complex pattern of neuropsychological deficits found in deep dyslexia. Hinton and Shallice [66••] trained an attractor network to settle into the appropriate distributed semantic representation of each of a set of words when presented with its written form. They found that lesions to the network (i.e., removing units or connections) resulted in both semantic errors (e.g., CAT $\Rightarrow$ 'dog') and visual errors (e.g., CAT $\Rightarrow$ 'cot') similar to that found in brain-injured patients with 'deep dyslexia,' obviating the need to posit separate lesions to account for each error type (cf. [67]). More recently, Plaut and Shallice [68•,69•] have extended the approach into a comprehensive account of the complex but repeatable combination of symptoms exhibited by these patients, and have demonstrated its generality over changes in network architecture, learning procedure, task definition, and output procedure. The same general approach has also been used to account for the perseverative and semantic effects on the object naming errors of optic aphasic patients [70•], and the degree of recovery and generalization in cognitive rehabilitation studies with acquired dyslexic patients [71•]. Some of the same findings can be accounted for in interactive localist models. Thus, Martin, Saffran, Dell, and Schwartz [72•] reproduced the semantic and phonological errors in naming and repetition of a deep dysphasic patient [73,74] by abnormally increasing the activity decay within Dell's [33] localist model of speech production. Furthermore, gradually reducing the decay parameter back towards its normal value mimics the patient's pattern of recovery.

Finally, we consider the application of connectionist models to the task of accounting for thought disorder in schizophrenia, in terms of an underlying disturbance in the regulation of the neuromodulator dopamine. Schizophrenics exhibit an abnormal (insufficient) use of context information in a variety of cognitive tasks, resulting from a disturbance in a mesocortical system that regulates dopamine levels in prefrontal cortex. Cohen and Servan-Schreiber [75••] showed that the neuromodulatory effects of dopamine on individual neurons can be well-approximated as a change in the input 'gain' of connectionist units, and that, over a number of tasks, reduction of gain in a pool of

units corresponding to prefrontal cortex produces the same behavioral deficits as exhibited by the patients. Thus, the modeling provides a clear computational link (gain) between the biological disturbance (abnormal dopamine levels) and the behavioral deficits (abnormal use of context) found in schizophrenia.

Conclusion

The connectionist framework for modeling human cognition has led to the development of explicit computational models of a wide range of cognitive functions. In many cases, these models introduce new ways of thinking about the nature of the computations that are performed and of the kind of learning that gives rise to the ability to carry out these computations. These models also give us new ways of thinking about the possible interpretations of psychological evidence, such as, for example, various psychological and neuropsychological dissociations. Based on this, it is becoming increasingly clear that connectionist models will play an important role in the further development of cognitive theory and in the further elaboration of the implications of phenomena observed in cognitive tasks.

Acknowledgments

We wish to thank Marlene Behrmann and Sue Becker for helpful comments on an earlier draft. Financial support is provided by the National Institute of Mental Health (Grants MH00385 and MH47566), the McDonnell-Pew Program in Cognitive Neuroscience (Grant T89- 01245-016) and the National Science Foundation (Grant ASC- 9109215).

References and Recommended Reading

1. Anderson JA, Silverstein JW, Ritz SA, Jones RS: Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model. Psych Rev 1977, 84:413–451.

2. McClelland JL: On the Time Relations of Mental Processes: An Examination of Systems of Processes in Cascade. Psych Rev 1979, 86:287–330.

3. McClelland JL, Rumelhart DE: An Interactive Activation Model of Context Effects in Letter Perception, Part I: An Account of Basic Findings. Psych Rev 1981, 88:375–407.

4. Rumelhart DE, McClelland JL, PDP Research Group: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations. Cambridge, MA: MIT Press; 1986.

5. McClelland JL, Rumelhart DE, PDP Research Group: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models. Cambridge, MA: MIT Press; 1986.

6. Hinton GE, McClelland JL and Rumelhart DE: Distributed Representations. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1. Edited by Rumelhart DE, McClelland JL, PDP research group. Cambridge, MA: MIT Press. 1986.

7. Sejnowski TJ, Koch C, Churchland PS: Computational Neuroscience. Science 1988, 241: 1299–1306.

8. Rumelhart DE, Hinton GE, Williams RJ: Learning Internal Representations by Error Propagation. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume I. Edited by Rumelhart DE, McClelland JL, PDP Research Group. Cambridge, MA: MIT Press; 1986.

9. Ackley DH, Hinton GE, Sejnowski TJ: A Learning Algorithm for Boltzmann Machines. Cogn Sci 1985, 9:147–169.

••10. Jordan MI, Rumelhart DE: Forward Models: Supervised Learning with a Distal Teacher. Cogn Sci 1992, 16:307–354. [Introduces a technique that dramatically extends the power of back-propagation to train neural networks. Give the extensions offered by Jordan and Rumelhart, it is possible to see how a neural network can be trained with error signals given in terms of mismatches between intended and obtained outcomes produced by outputs, rather than simply intended and obtained outputs themselves.]

11. Linsker R: An Application of the Principle of Maximum Information Preservation to Linear Systems. In Advances in Neural Information Processing Systems 1. Edited by Touretzky, DS. San Mateo, CA: Morgan Kaufmann; 1989:186–194.

•12. Weigand AS, Rumelhart DE, Huberman BA: Generalization by Weight-Elimination with Application to Forecasting. In Advances in Neural Information Processing Systems 3. Edited by Lippman RP, Moody JE, Touretzky DS. San Mateo, CA: Morgan Kaufmann; 1991. [Introduces a widely-used method for improving generalization in connectionist networks through forcing the network to find a solution that minimizes the number of connection weights used.]

13. Jacobs RA, Jordan MI: Computational Consequences of a Bias toward Short Connections. J Cogn Neurosci 1992, 4:323–336.

14. Miller KD, Keller JB, Stryker MP: Ocular Dominance Column Development: Analysis and Simulation. Science 1989,

•15. Miller KD: Development of Orientation Columns via Competition between ON- and OFF-Center Inputs. NeuroReport 1992, 3:73–76. [Suggests how a key aspect of the visual system may arise from simple developmental processes.]

•16. Jacobs RA, Jordan MI, Barto AG: Task Decomposition through Competition in a Modular Connectionist Architecture: The What and Where Vision Tasks. Cogn Sci 1991, 15:219–250. [Provides a mechanism whereby networks can modularize themselves, demonstrating how this can lead to improved learning and generalization.]

•17. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE: Adaptive Mixtures of Local Experts. Neural Computation 1991, 3:79–87. [Refines the mechanism proposed by [•16].]

18. McClelland JL: Toward a Theory of Information Processing in Graded, Random, Interactive Networks. In Attention & Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience. Edited by Meyer DE, Kornblum S. Cambridge, MA: MIT Press; in press.

19. Massaro DW: Testing between the TRACE Model and the Fuzzy Logical Model of Speech Perception. Cogn Psychol 1989, 21:398–412.

•20. McClelland JL: Stochastic Interactive Processes and the Effect of Context on Perception. Cogn Psychol 1991, 23:1–44. [Shows that data thought to be problematic for interactive models are actually fully consistent with stochastic interactive models, though not deterministic interactive ones.]

21. Rumelhart DE, McClelland JM: On Learning the Past Tenses of English Verbs. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2. Edited by McClelland JL, Rumelhart DE, PDP Research Group. Cambridge, MA: MIT Press; 1986.

22. Plunkett K, Marchman VA: U-Shaped Learning and Frequency Effects in a Multi-Layered Perceptron: Implications for Child Language Acquisition. Cognition 1991, 38:43–102.

23. MacWhinney B, Leinbach J: Implementations are not Conceptualizations: Revising the Verb Learning Model. Cognition 1991, 40:121–153.

24. Daugherty K, Seidenberg MS: Rules or Connections? The Past Tense Revisited. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1992: 259–264.

25. Hoeffner J: Are Rules a Thing of the Past? The Acquisition of Verbal Morphology by an Attractor Network. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1992: 861–866.

•26. Hare M, Elman JL: A Connectionist Account of English Inflectional Morphology: Evidence from Language Change. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1992: 265–270. [Shows how a connectionist network can capture the imposition of structure on a language system by the learner and the evolution of languages through successive generations of learners.]

••27. Elman JL: Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. Machine Learning 1991, 7:195- -225. [Demonstrates that simple recurrent networks can recover a key aspect of the structure of natural languages, including long- distance dependencies that span embedded clauses.]

28. Pinker S: Rules of Language. Science 1991, 253:530–535.

29. Miikkulainen R, Dyer MG: Natural Language Processing with Modular PDP Networks and Distributed Lexicon. Cogn Sci 1991, 15:343–399.

30. Gupta P, MacWhinney B: Integrating Category Acquisition with Inflectional Marking: A Model of the German Nominal System. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science

Society. Hillsdale, NJ: Erlbaum; 1992:253–258.

31. Levelt WJM, Schriefers H, Vorberg D, Meyer AS, Pechmann T, Havinga J: The Time Course of Lexical Access in Speech Production: A Study of Picture Naming. Psychol Rev 1991, 98:122–142.

•32. Dell GS, O'Seaghdha PG: Stages of Lexical Access in Language Production. Cognition, 1992, 42:287–314. [An account of data [31] suggesting discrete states in speech production within a localist, spreading-activation connectionist system [33] which is globally modular but locally interactive.]

33. Dell GS: A Spreading-Activation Theory of Retrieval in Sentence Production. Psychol Rev 1986, 93:283–321.

•34. Dell GS, Juliano C, Govindjee A: Structure and Content in Language Production: A Theory of Frame Constraints in Phonological Speech Errors. Cogn Sci 1993, in press. [A systematic analysis of the ability of distributed networks to learn the statistical structure of English word pronunciations (as evidenced by patterns of speech errors) without built-in structure.]

•35. McClelland JL: The Interaction of Nature and Nurture in Development: A Parallel Distributed Processing Perspective. To appear in a forthcoming book edited by Richelle, Eelen, Bertelson and d'Ydewalle. Hillsdale, NJ: Erlbaum. [Re-examines evidence and arguments for nativist approaches to development in light of connectionist models.]

36. Bates EA, Elman JL: Connectionism and the Study of Change. In Brain Development and Cognition. Edited by Johnson MH. Oxford: Blackwell; in press.

37. Karmiloff-Smith A: Beyond Modularity: A Developmental Perspective on Cognitive Science. Cambridge, MA: MIT Press; 1992.

38. Plunkett K, Sinha C: Connectionism and Developmental Theory. Psykologisk Skriftserie Aarhus 1991, 16:1–34.

39. Newport EL: Maturational Constraints on Language Learning. Cogn Sci 1990, 14:11–28.

••40. Elman JL: Incremental Learning, or the Importance of Starting Small. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1991: 443–448. [See next.]

••41. Elman JL: Learning and Development in Neural Networks: The Importance of Starting Small. Cognition, submitted. [Simple recurrent networks succeed at learning a difficult, English-like grammar only when the computational resources of the network are initially limited and gradually increased. Provides a specific explanation for 'critical period' effects in language learning in terms of how immaturity compensates for limitations in the learning abilities of connectionist networks.]

••42. Marchman VA: Language Learning in Children and Neural Networks: Plasticity, Capacity and the Critical Period. J Cogn Neurosci 1993, in press.

[Replicates many critical period effects in networks which are lesioned at various points during learning the English past-tense. Argues that poorer learning and recover from damage later in life may results from the inflexibility that results as knowledge becomes 'entrenched' in connection weights.]

43. Markman EM: Constraints Children Place on Word Meanings. Cogn Sci 1990, 14:57–77.

•44. Schyns PG: A Modular Neural Network Model of Concept Acquisition. Cogn Sci 1991, 15:461–508. [Presents simulations in which lower-level feature-based representations of concepts are influences by the top-down pressures from naming, accounting for a number of aspects of children's learning of words for concepts.]

45. Cleeremans A, McClelland JL: Learning the Structure of Event Sequences. J Exp Psychol [Gen] 1991, 120:235–253.

46. McCloskey M, Cohen NJ: Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In The Psychology of Learning and Motivation: Advances in Research and Theory, 24. Edited by Bower GH. San Diego: Academic Press; 1989: 109–165.

47. Kruschke JK: ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. Psychol Rev 1992, 99:22–44.

•48. McClelland JL, McNaughton BL, O'Reilly R, Nadel L: Complementary Roles of Hippocampus and Neocortex in Learning and Memory. Society for Neuroscience Abstracts 1992, 18(508.7):1216.

•49. Schmajuk NA and DiCarlo JJ: Stimulus Configuration, Classical Conditioning, and Hippocampal Function. Psychol Rev 1992, 99:268- -305. [Presents a detailed account of effects of the role of the Hippocampus in classical conditioning, accounting for the special role of the hippocampal formation in configural cue learning.]

50. Gluck MA, Meyers CE: Hippocampal-System Function in Stimulus Representation and Generalization: A Computational Theory. In Proceedings of the Fourteenth Annual Conference of the Cognitive

Science Society. Hillsdale, NJ: Erlbaum; 1992: 390–395.

•51. Olshausen B, Anderson C, Van Essen D: A Neural Model of Visual Attention and Invariant Pattern Recognition [Technical Report]. Pasadena, CA: California Institute of Technology, Computation and Neural Systems Program; 1992: CNS Memo 18. [A model of visual attention in which control signal route the cortical activity caused by an object anywhere in the visual field onto the same higher-level units for recognition via template-matching. Attempts to draw very explicit correspondences between structures in the model and specific visual cortical areas.]

••52. Hummel JE, Biederman I: Dynamic Binding in a Neural Network for Shape Recognition. Psychol Rev 1992, 99:480–517. [A detailed proposal for viewpoint-invariant shape recognition using temporal synchrony to segment the visual scene into geometric primitives (geons) that together form a structural description of an object. Relates to both neurophysiological and psychological data on intermediate- and high-level vision.]

53. Biederman I: Recognition-by-Components: A Theory of Human Image Understanding. Psychol Rev 1987, 94:115–147.

54. Kosslyn SM, Chabris CF, Marsolek CJ, Koenig O: Categorical versus Coordinate Spatial Relations: Computational Analyses and Computer Simulations. J Exp Psychol [Hum Percept] 1992, 18:562– 577.

55. Gray CM, Konig P, Engel AK, Singer W: Oscillatory Responses in Cat Visual Cortex Exhibit Inter-Columnar Synchronization which Reflect Global Stimulus Properties. Nature 1989, 338:334–337.

56. Mozer MC, Zemel RS, Behrmann M, Williams CKI: Learning to Segment Images using Dynamic Feature Binding. Neural Computation 1992, 4:650–665.

•57. Shastri L, Ajjanagadde V: From Simple Associations to Systematic Reasoning: A Connectionist Representation of Rules, Variables, and Dynamic Bindings using Temporal Synchrony. Behav Brain Sci 1993, in press. [A detailed proposal for binding concepts and roles using temporal synchrony during implicit reasoning, within a structured connectionist framework.]

•58. Smolensky P, Legendre G, Miyata Y: Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition [Technical Report]. Boulder, CO: University of Colorado at Boulder, Computer Science Department and Institute of Cognitive Science; 1992: CU-CS-600-92. [A thorough treatment of a variety of methodological issues in cognitive science, arguing for the use of connectionist modeling as a unifying formalism. A specific tensor- product formalism is developed and applied to problems in linguistics and philosophy of mind. A definition of well- formedness, or 'Harmony' applies to mental processes both as massively parallel spreading activation, and as manipulation of symbol structures.]

59. Fodor JA, Pylyshyn ZW: Connectionism and Cognitive Architecture: A Critical Analysis. Cognition 1988, 28:3–71.

60. Posner MI, Walker JA, Friedrich FJ, Rafal RD: Effects of Parietal Injury on Covert Orienting of Visual Attention. J Neurosci 1984, 4:1863–1874.

•61. Cohen JD, Romero RD, Servan-Schreiber D, Farah MJ: Disengaging from the Disengage Function: The Relation of Macrostructure to Microstructure in Parietal Attentional Deficits. J Cogn Neurosci, submitted. [Damage within a competitive architecture for allocating attention can account for data suggesting impairment to a dedicated 'disengage' mechanism.]

62. Warrington EK, Shallice T: Category Specific Semantic Impairments. Brain 1984, 107:829–853.

•63. Farah MJ, McClelland JL: A Computational Model of Semantic Memory Impairment: Modality-Specificity and Emergent Category- Specificity. J Exp Psychol [Gen] 1991, 120:339–357. [Following [62], demonstrates that apparent category-specific deficits (living vs. nonliving things) can arise from a modality-specific deficit (visual vs. functional semantics). Goes on to show how interactivity between visual and functional semantics can account for why damage to visual semantics impairs functional knowledge of living things.]

•64. Burton MA, Young AW, Bruce V, Johnston RA, Ellis AW: Understanding Covert Recognition. Cognition 1991, 39:129–166. [A dissociation between overt and covert face recognition in prosopagnosia can be produced in a localist model of face processing by reducing the support that identification nodes (needed for naming) receive from face recognition nodes.]

•65. Farah MJ, O'Reilly RC, Vecera SP: Dissociated Overt and Covert Recognition as an Emergent Property of Lesioned Attractor Networks. Psychol Rev, submitted. [An account of the dissociation between overt and covert face recognition in terms of residual knowledge in a partially-damaged attractor network that associates names and faces via semantics.]

••66. Hinton GE, Shallice T: Lesioning an Attractor Network: Investigations of Acquired Dyslexia. Psychol Rev 1991, 98:74–95. [Lesions to an attractor network that maps orthography to semantics reproduce

the co-occurrence of semantic errors and visual errors found in deep dyslexia. Demonstrates how principles from connectionist modeling, such as distributed representations and attractors, can provide insight into what otherwise are perplexing phenomena.]

67. Coltheart M, Patterson KE, Marshall JC: Deep Dyslexia. London: Routledge & Kegan Paul; 1980.

•68. Plaut DC, Shallice T: Effects of Word Abstractness in a Connectionist Model of Deep Dyslexia. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1991: 73–78. [Extends the approach in [66••] to account for the effects of abstractness and their interactions with visual similarity in the oral reading errors made by deep dyslexic patients.]

•69. Plaut DC, Shallice T: Deep Dyslexia: A Case Study of Connectionist Neuropsychology. Cogn Neuropsychol 1993, in press. [Presents a systematic analysis of a range of network architectures, learning procedures, task definitions, and output interpretation procedures, aimed at evaluating and extending the empirical adequacy and computational generality of the attractor network account of deep dyslexia [66••].]

•70. Plaut DC, Shallice T: Perseverative and Semantic Influences on Visual Object Naming Errors in Optic Aphasia: A Connectionist Account. J Cogn Neurosci 1993, 5:89–116. [An attractor network augmented with short-term correlational weights is trained to generate semantic representations of objects from high-level visual representations. Under damage, the network exhibits the complex semantic and perseverative effects of patients with a visual naming disorder known as 'optic aphasia.']

•71. Plaut DC: Relearning after Damage in Connectionist Networks: Implications for Patient Rehabilitation. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1992: 372–377. [An attractor network that maps orthography to semantics is retrained after damage. Recovery and generalization is significant (like some acquired dyslexics) after lesions within semantics but not after lesions near orthography.]

•72. Martin N, Saffran EM, Dell GS, Schwartz MF: Origins of Paraphasias in Deep Dysphasia: Testing the Consequences of a Decay Impairment to an Interactive Spreading Activation Model of Lexical Retrieval. Cognition, submitted. [Increasing the decay in the activity of lexical units in Dell's [33] model of speech production produces semantic and phonological errors like a patient with deep dysphasia [73]. Gradually decreasing the decay back towards normal simulations the change in error pattern during the patient's recovery.]

73. Martin N, Saffran EM: Repetition and Verbal STM in Transcortical Sensory Aphasia: A Case Study. Brain Lang 1990, 39:254–288.

74. Martin N, Saffran EM: A Computational Account of Deep Dysphasia: Evidence from a Single Case Study. Brain Lang 1992, 43:240–274.

••75. Cohen JD, Servan-Schreiber D: Context, Cortex, and Dopamine: A Connectionist Approach to Behavior and Biology in Schizophrenia. Psychol Rev 1992, 99:45–77. [A unified computational interpretation of abnormalities in schizophrenia at both the neurochemical and behavioral levels. Abnormal levels of dopamine in prefrontal cortex lead to abnormal use of context information by schizophrenics in a variety of task. In analogous connectionist models, abnormal setting of a 'gain' parameter in the portion of each model corresponding to prefrontal cortex affects units in the same way as dopamine affects neurons, and also produces the same behavioral abnormalities.]

Address of Authors

James L. McClelland and David C. Plaut
Department of Psychology
Carnegie Mellon University
Pittsburgh PA 15213–3890
jlm@andrew.cmu.edu and plaut@cmu.edu