Commentary on Anderson and Lebiere

Abstract: *153* words
Main Text: *4005* words
References: *824* words
Total Text: *5332* words (includes above plus footnotes)

Developing a domain-general framework for cognition: What is the best approach?

James L. McClelland
Department of Psychology and Center for the Neural Basis of Cognition
Carnegie Mellon University
115 Mellon Institute
Pittsburgh, PA 15213
USA
+1-412-268-3157 (for correspondence)
jlm@cnbc.cmu.edu
http://www.cnbc.cmu.edu/~jlm

David C. Plaut
Department of Psychology and Center for the Neural Basis of Cognition
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
USA
plaut@cmu.edu
http://www.cnbc.cmu.edu/~plaut

Stephen J. Gotts
Department of Psychology and Center for the Neural Basis of Cognition
Carnegie Mellon University
and
Laboratory of Neuropsychology
NIMH/NIH
Building 49, Suite 1B-80
49 Convent Drive
Bethesda, MD 20892
USA
gotts@nih.gov
http://www.cnbc.cmu.edu/~gotts

Tiago V. Maia
Department of Psychology
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
USA

tmaia@andrew.cmu.edu
http://www.cnbc.cmu.edu/~tmaia

Abstract

We share with Anderson and Lebiere (and with Newell before them) the goal of developing a domain-general framework for modeling cognition, and we take seriously the issue of evaluation criteria. We advocate a more focused approach than the one reflected in Newell's criteria, based on analysis of failures as well as successes of models brought into close contact with experimental data. Anderson and Lebiere attribute the shortcomings of our parallel-distributed processing framework to a failure to acknowledge a symbolic level of thought. Our framework does acknowledge a symbolic level, contrary to their claim. What we deny is that the symbolic level is the level at which the principles of cognitive processing should be formulated. Models cast at a symbolic level are sometimes useful as high-level approximations of the underlying mechanisms of thought. The adequacy of this approximation will continue to increase as symbolic modelers continue to incorporate principles of parallel distributed processing.

In their target article, Anderson and Lebiere (A&L) present a set of criteria for evaluating models of cognition, and rate both their own ACT-R framework and what they call 'classical connectionism' on the criteria. The Parallel Distributed Processing (PDP) approach, first articulated in the two PDP volumes (Rumelhart, McClelland, & the PDP Research Group, 1986) appears to be close to the prototype of what they take to be 'classical connectionism'. While we cannot claim to speak for others, we hope that our position will be at least largely consistent with that of many others who have adopted connectionist/PDP models in their research.

There are three main points that we would like to make.

1. We share with A&L (and with Newell before them) the effort to develop an overall framework for modeling human cognition, based on a set of domain-general principles of broad applicability across a wide range of specific content areas.

2. We take a slightly different approach from the one that Newell advocated to pursuing the development of our framework. We think it worthwhile to articulate this approach briefly, and to comment on how it contrasts with the approach advocated by Newell and apparently endorsed by A&L.

3. We disagree with A&L's statement that classical connectionism denies a symbolic level of thought. What we deny is only the idea that the symbolic level is the level at which the principles of processing and learning should be formulated. We treat symbolic cognition as an emergent phenomenon that can sometimes be approximated by symbolic models, especially those that incorporate the principles of connectionist models.

In what follows, we elaborate these three points, addressing the first one only briefly, since this is a point of agreement between us and A&L.

The Search for Domain-General Principles

There is a long-standing tradition within psychological research to search for general principles that can be used to address all aspects of behavior and cognition. With the emergence of computational approaches in the 1950's and 1960's, and with the triumph of the von Neumann architecture as the basis for artificial computing devices, this search could be formulated as an effort to propose what Newell called "a unified architecture for cognition". An architecture consists of a specification of (a) the nature of the building blocks out of which representations and processes are constructed, (b) the fundamental rules by which the processes operate, and (c) an overall organizational plan that allows the system as a whole to operate. Newell's SOAR architecture and A&L's ACT-R architecture are both good examples of architectures of this type. For our part, we have sought primarily to understand (a) the building blocks and (b) the fundamental rules of processing. Less effort has been devoted to the specifics of the overall organizational plan as such, although we do take a position on some of the principles that the organizational plan instantiates. Because the organization is not fully specified as such, we find it more congenial to describe what we are developing as a framework than an architecture. But this is a minor matter; the important point is the shared search for general principles of cognition.

We are of course well aware that this search for general principles runs counter to a strong alternative thread that treats distinct domains of cognition as distinct cognitive modules that operate according to domain-specific principles. Such a view has been articulated for language by Chomsky, for Vision by Marr. Fodor and Keil have argued the more general case, and a great deal of work has been done to try to elucidate the specific principles relevant to a wide range of alternative domains. While we cannot prove that this approach is misguided, we have the perspective that the underlying machinery and the principles by which it operates is fundamentally the same across all different domains of cognition. While this machinery can be tuned and parameterized for domain-specific uses, understanding the broad principles by which it operates will necessarily be of very broad relevance.

How the Search for Domain General Principles is Carried Out

If one's goal is to discover the set of domain general principles that govern all aspects of human cognition, how best is the search for such principles carried out? Our approach begins with the fundamental assumption that it is not possible to know in advance what the right set of principles are. Instead, something like the following discovery procedure is required.

a. Begin by formulating a putative set of principles.

b. Develop models based on these principles and apply them to particular target domains (i.e., bodies of related empirical phenomena).

c. Assess the adequacy of the models so developed, and attempt to understand what really underlies both successes and failures of the models.

d. Use the analysis to refine and elaborate the set of principles, and return to step b.

In practice this appears to be the approach both of Newell and of A&L. Newell and his associates developed a succession of cognitive architectures, as has Anderson; indeed Newell suggested that his was only really one attempt, and that others should put forward their own efforts. However, Newell argued for broad application of the framework across all domains of cognition, suggesting that an approximate account within each would be satisfactory. In contrast, we advocate a more focused exploration of a few informative target domains, and using failures of proposed models to guide further explorations of how the putative set of principles should be elaborated. To illustrate the power of this approach, we briefly review two cases. Note that we do not mean to suggest that Anderson and Lebiere

explicitly advocate the development of approximate accounts. Rather, our point is to bring out the importance of focus in bringing out important principles of cognition.

1. The interactive activation model (McClelland & Rumelhart, 1981) explored the idea that context effects in perception of letters--- specifically, the advantage for letters in words relative to single letters in isolation---could be attributed to the bi-directional propagation of excitatory and inhibitory signals among simple processing units whose activation corresponds to the combined support for the item the unit represents. When a letter occurs in a word, it and the other letters will jointly activate the unit for the word, and that unit will in turn send additional activation back to each of the letters, thereby increasing the probability of recognition. Similar ideas were later used in the TRACE model of speech perception (McClelland and Elman, 1986) to account for lexical influences on phoneme identification. Massaro (1989; Massaro & Cohen, 1991) pointed out that the interactive activation model failed to account for the particular quantitative form of the influence of context on the identification of a target item. He argued that the source of the problem lay specifically in the use of bi-directional or interactive activation between phoneme or letter units on the one hand and word units on the other. Since the interactive activation model fit the data pretty well, Newell might have advocated accepting the approximation, and moving on to other issues. However, close investigation of the issue turned out to lead to an important discovery. Subsequent analysis (McClelland, 1991; Movellan & McClelland, 2001) showed that the failure of the interactive activation model arose from faulty assumptions about the source of variability in performance. Discovering this was made possible by the failure of the model. It then became possible to ask, what changes had to be made in order to fit the data? McClelland (1991) showed that the model had a general deficiency in capturing the joint effects of two different sources of influence even if they were both bottom up and activation was only allowed to propagate in a feed-forward direction. The problem was attributed instead to the fact that in the original McClelland and Rumelhart model, the interactive activation process was completely deterministic and activations were transformed into response probabilities only at the moment of response selection. This led to the discovery of what we take to be an important principle: that the activation process is not only graded and interactive but also intrinsically variable. Reformulated versions of the model incorporating intrinsic variability in addition to gradedness and interactivity were shown through simulations (McClelland, 1991) and mathematical analysis (Movellan & McClelland, 2001) to produce the right quantitative form of contextual influence on phoneme and letter identification. This principle of intrinsic variability has been incorporated in several subsequent models, including a model that addresses in detail the shapes of reaction time distributions and the effects of a variety of factors on these distributions (Usher & McClelland, 2001).

2. Seidenberg and McClelland (1989) introduced a model of single word reading that accounted for frequency, regularity, and consistency effects in single word reading. The model relied on a single network that mapped distributed input representations of the spellings of words, via one layer of hidden units, onto a set of output units representing the phonemes in the word's pronunciation. However, as two independent critiques pointed out (Besner, Twilley, McCann, & Seergobin, 1990; Coltheart, Curtis, Atkins, & Haller, 1993) the model performed far worse than normal human subjects at reading pronounceable nonwords. Both critiques attributed this shortcoming of the model to the fact that it did not rely on separate lexical and rule-based mechanisms. However, subsequent connectionist research (Plaut, McClelland, & Seidenberg, 1995; Plaut, McClelland, Seidenberg, & Patterson, 1996) demonstrated that the particular choice of input and output representations used by Seidenberg and McClelland (1989) was instead the source of the difficulty. These representations tended to disperse the regularity in the mapping from spelling to sound over a number of different processing units. This was because the input units activated by a given letter depended on the surrounding context, and the output units representing a given phoneme were likewise context dependent. Since the learning in the model is in the connections among the units, this led to a dispersion of the information about the regularities across many different connections, and created a situation in which letters in nonwords might occur in

contexts that had not previously been encountered by the network. This led to the discovery of the principle that to succeed in capturing human levels of generalization performance, the representations used in connectionist networks must condense the regularities. Subsequent models of word reading, inflectional morphology, and other cognitive tasks have used representations that condense the regularities, leading them to achieve human levels of performance with novel items while yet being able to learn to process both regular and exception words. (Footnote 1)

These two case studies bring out the importance of taking seriously mismatches between a model's behavior and human performance data, even when the model provides an approximate account of most of the relevant phenomena. We believe that such mismatches are important forces in driving the further development of a framework. Of course such mismatches might also reflect a fundamental inadequacy of the framework as a whole or of its most fundamental grounding assumptions. Analysis is required to determine which; but whatever the outcome, the examination of failures of fit is an important source of constraint on the further development of the framework.

With these comments in mind, we can now turn to the framing of the goals of cognitive modeling as articulated in the sorts of criteria that Newell proposed and A&L have adopted with their own modifications. We agree that it is useful to focus attention on some of these general issues, and that that there is more to a good cognitive model than simply a close fit to experimental data. We would note, however, that making the effort at this stage to achieve the sort of breadth that Newell's criteria imply may distract attention from addressing critical discrepancies that can only be revealed through close comparison of models and data. We have chosen to adopt a more focused approach, but we do not deny that a broader approach may reveal other limitations, and that it may be worthwhile for some researchers to follow Newell's strategy.

The Importance and Nature of the Symbolic Level

A&L suggest that the shortcomings of the connectionist approach are fundamental, deriving from its failure to acknowledge a symbolic level of thought, whereas the shortcomings of the ACT-R theory are temporary, and derive from its failure as yet to address certain of Newell's criteria. We have a very different reading of the situation.

First of all, our Parallel Distributed Processing approach does not deny a symbolic level of thought. What we deny is only that the symbolic level is the appropriate level at which the principles of processing and learning should be formulated. We treat symbolic thought as an emergent phenomenon which can sometimes be approximated to a degree by a model formulated at the symbolic level, but which, on close scrutiny, does not conform exactly to the properties that it should have according to symbolic models.

As is well known, the issue here is one that has been extensively explored in the context of research on the formation of past tenses and other inflections of nouns and verbs. A recent exchange of articles contrasts the PDP perspective (McClelland & Patterson, 2002a, 2002b) and Pinker's symbolic, dual-mechanism account (Pinker & Ullman, 2002b, 2002a). Here we will present the PDP perspective.

In several places, Pinker and his colleagues have argued that the past tense of English is characterized by two mechanisms, one involving symbolic rules and the other involving a lexical mechanism that operates according to connectionist principles. A symbolic rule, according to Pinker's approach, is one that applies uniformly to all items that satisfy its conditions. Furthermore, such conditions are abstract and very general. For example, the past tense rule applies uniformly to any string of phonemes, provided only that it is the stem of a verb. In many places Pinker also states that symbolic rules are acquired

suddenly; this conforms to the idea that a rule is something that one either has or does not have. Finally, the symbolic rule is thought to require a completely different kind of mechanism than the one underlying the inflection of exceptions, leading to the prediction that brain lesions could selectively impair the ability to use the rule while leaving the inflection of irregular forms intact.

Although Pinker and his colleagues have pointed to evidence they believe supports their characterization of the mechanism that produces regular past-tense inflections, in their review of that evidence McClelland and Patterson (2002) found instead that in every case the evidence supports an alternative characterization, first proposed by Rumelhart and McClelland (1986), in which the formation of an inflected form arises from the interactions of simple processing units via weighted connections learned gradually from exposure to example forms in the language (Footnote 2). Specifically, the evidence indicates that the onset of use of regular forms is gradual (extending over a full year, Brown, 1973; Hoeffner, 1996). It is initially restricted to verbs characterized by a set of shared semantic properties, and then gradually spreads to other verbs starting with those sharing some of the semantic properties of the members of the initial set (Shirai & Anderson, 1995). Usage of the regular past tense by adults is not insensitive to phonology but instead reflects phonological and semantic similarity to known regular verbs (Albright & Hayes, 2001; Ramscar, 2002). Furthermore, purported dissociations arising from genetic defects (Gopnik & Crago, 1991) or strokes (Ullman et al., 1997) disappear when materials are used that control for frequency and phonological complexity (Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995; Bird, Lambon Ralph, Seidenberg, McClelland, & Patterson, 2003); individuals with deficits in inflection of regular forms show corresponding deficits with appropriately matched exceptions. In short, the acquisition and adult use of the regular past tense exhibits exactly those characteristics expected from the connectionist formulation. Ultimate adult performance on regular items conforms approximately to the predictions of the rule; for example, reaction time and accuracy inflecting regular forms is relatively insensitive to the word's own frequency. But exactly the same effect also arises in the connectionist models; as they learn from many examples that embody the regular pattern, the connection weights come to reflect it in a way that supports generalization to novel items and makes the number of exposures to the item itself relatively unimportant.

In summary, the characteristics expected on a connectionist approach, but not the symbolic rule approach of Pinker, are exhibited by human performance in forming inflections. Such characteristics include fairly close approximation to what would be expected from use of a symbolic rule under specifiable conditions, but allow for larger discrepancies from what would be predicted from the rule under other conditions (i.e., early in development, after brain damage of particular kinds, and when the language environment is less systematic). (Footnote 3)

What implications do the characteristics of human performance in forming inflections have for the ACT-R approach of A&L? They have already described an ACT-R model (Taatgen & Anderson, in press) of past tense formation in which the acquisition of the regular past tense occurs fairly gradually, and we have no doubt that with adjustment of parameters even more gradual acquisition would occur. Furthermore, we see relatively little in A&L's formulation that ties them to the claim that the conditions for application of symbolic rules must be abstract that Pinker (1991, Pinker & Ullman, 2002a) and Marcus (2001) have made. Nor is there anything that requires them to posit dissociations, since production rules are used in their model for both regular and exceptional forms. Thus, although the past tense rule actually acquired in the Taatgen and Anderson model is as abstract and general as the one proposed by Pinker, a modified version of their model could surely be constructed bringing it closer to the connectionist account. To capture the graded and stochastic aspects of human performance, they have introduced graded strengths that are tacked onto symbolic constructs (propositions and productions), thereby allowing them to capture graded familiarity and regularity effects. To capture similarity effects, there is no reason why the condition-matching operation performed by rule-like productions could not be formulated as graded constraints, so that the degree of activation of a

production would depend on the degree to which its conditions match current inputs. Indeed Anderson and Lebiere note that by allowing graded condition-matching in ACT-R they can capture the graded, similarity-based aspects of human performance that are naturally captured within the connectionist framework.

Even these adjustments, however, would leave one aspect of connectionist models unimplemented in the Taatgen and Anderson model. This is the ability of connectionist models to exploit multiple influences simultaneously, rather than depending on the output generated by just one production at a time. Specifically, in the Taatgen and Anderson account of past-tense formation, a past tense form is generated either by the application of the general ed-rule or by the application of an item-specific production; the form that is generated depends on only one of these productions, and not on their simultaneous activation. We argue that this is a serious weakness, in that it prevents the Taatgen and Anderson model from exploiting the high degree of conformity with the regular pattern that exists among the exceptions. In our view this is an important and general limitation of many symbolic models, even ones like ACT-R that have moved a long way toward incorporating many of the principles of processing espoused by connectionists.

As McClelland and Patterson (2002b) have noted, fully 59% of the exceptional past tense verbs in English end in /d/ or /t/. In the connectionist models, the same connection-based knowledge that imposes the regular inflection on fully regular verbs also operates in the inflection of these exceptional cases. That is, the same connections that add /t/ to regular 'like' to make 'liked' also add /t/ to irregular 'keep' to make 'kept'. In the case of kept, additional influences (from experience with kept itself and other similar cases) also operate to allow the model to capture the alteration of the vowel that makes this item an exception. In contrast, in the Taatgen and Anderson model and many other dual-mechanism models, only one production at a time can fire, so that a past tense form is either generated by the rule (in which case it will be treated as regular) or by a production specific to it as an exception. Given this, no benefit accrues to an exception for sharing properties of the regular past tense, and all exceptions might as well be completely arbitrary. This is problematic because it leaves unexplained important aspects of the distributions of word forms. Across languages, there are many forms that are partially regular and very few that are completely arbitrary, and those that are completely arbitrary are of very high frequency (Plunkett & Marchman, 1991; and the same is true for irregular spelling-to-sound correspondences). This suggests that human language users are highly sensitive to the degree to which exceptions share properties with regular items, contrary to the properties of the Taatgen and Anderson model.

In response to this, we anticipate that A&L might be tempted to modify the ACT-R framework even further in the direction of connectionist models by allowing application of multiple productions to work together to produce an individual inflected word form. We certainly think this would lead to models that would be more likely than current ACT-R based accounts to address the influence of regularities in exceptions, and would bring ACT-R more fully into like with the fundamental idea of parallel distributed processing. After all the essence of PDP is the idea that every act of cognition depends on and is distributed over a large number of contributing units, quite different from what happens presently in ACT-R, where any given output is the product of the application of a single production.

While such a change to ACT-R would, we believe, improve it considerably, we would simply note two points in this context. First, this would continue the evolution of symbolic models of human cognition even further in a connectionist-like direction. This evolution, which has been in process for some time, is not, in our view, accidental, since with each step in this direction, symbolic models have achieved a higher degree of fidelity to the actual properties of human cognition. What this indicates to us is that, while the shortcomings of symbolic models may be temporary as A&L suppose, they are most likely to be overcome by incorporation of the very principles that govern processing as defined at the connectionist level.

Second, as symbolic modelers take each new step in the direction of connectionist models, they do so in acceptance of the fact that the phenomena to be explained have the characteristics that served to motivate the exploration of connectionist models in the first place. This, in turn, undermines the stance that the fundamental principles of human cognition should be formulated at the symbolic level, and instead further motivates the exploration of principles at the connectionist level. While we acknowledge that connectionist models still have many limitations, we nevertheless feel that this does not arise from any failure to acknowledge a symbolic level of thought. Instead we suggest it arises from the fact the connectionists (like symbolic modelers) have not yet had the chance to address all aspects of cognition or all factors that may affect it.

In spite of our feeling that the facts of human cognition are completely consistent with the principles of parallel distributed processing, we do not wish to give the impression that we see no merit in modeling that is directed at the symbolic level. Given that symbolic formulations often do provide fairly good approximations, it may be useful to employ them in cases where it is helpful to exploit their greater degree of abstraction and succinctness. We feel that work at a symbolic level will proceed most effectively if it is understood that it approximates a system that is underlyingly much more parallel and distributed, since at that point insights from work at the connectionist level will flow even more freely into efforts to capture aspects of cognition at the symbolic level.

Footnotes:

1. It is necessary to note that none of the models we have discussed fully embody all of the principles of the PDP framework. For example, the interactive activation and TRACE models use localist, not distributed, representations, while the models of spelling-to-sound mapping (Seidenberg & McClelland, 1989; Plaut et al, 1996) do not incorporate intrinsic variability. This fact can lead to confusion about whether indeed there is a theoretical commitment to a common set of principles. In fact we do have such a commitment. The fact that individual models do not conform to all of the principles is a matter of simplification. This leads to computational tractability and can foster understanding, and we only adopt these practices for these reasons. Everyone should be aware that models that are simplified embodiments of the theory do not demonstrate that models incorporating all of its complexity will be successful. In such cases further research is necessary, especially when the possibility of success is controversial. For example, Joanisse and Seidenberg (1999) used localist word units in their model of past-tense inflection, and Pinker and Ullman (2002a, 2002b) have argued that this is essential. In this context, we fully accept that further work is necessary to demonstrate that a model using distributed semantic representations can actually account for the data.

2. It should be noted here that none of these models assume that learning occurs through correction of overtly generated errors. Instead it is assumed that exposure provides examples of appropriate usage in context. The learner uses the context as input to generate an internal representation corresponding to the expected phonological form. Learning is driven by the discrepancy between this internal representation and the actual perceived form provided by the example.

3. Marcus et al claimed that German has a regular plural (the so-called +s plural) that conforms to the expectation of the symbolic approach, in spite of the fact that it is relatively infrequent. However, subsequent investigations indicate that the +s plural does not exhibit the properties one would expect if it were based on a symbolic rule (Bybee, 1995; Hahn & Nakisa, 2000).

References

Albright, A., & Hayes, B. (2001, November). Rules vs. analogy in English past tenses: A computational/experimental study. (Department of Linguistics. UCLA).

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? Psychological Review, 97(3), 432-446.

Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? Neuropsychologia, 48, 502-526.

Brown, R. (1973). A first language. Cambridge, MA: Harvard University Press.

Bybee, J. L. (1995) Regular morphology and the lexicon. Language and Cognitive Processes, 10, 425-455.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological Review, 100(4), 589-608.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. Cognition, 39, 1-50.

Hahn, U & Nakisa, R. C. (2000). German inflection: Single-route or dual-route? Cognitive Psychology, 41, 313-360.

Hoeffner, J. H. (1996). A single mechanism account of the acquisition and processing of regular and irregular inflectional morphology. Unpublished doctoral dissertation, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.

Joanisse, M.F., & Seidenberg, M.S. (1999). Impairments in verb morphology following brain injury: a connectionist model. Proceedings of the National Academy of Sciences, 96, 7592-7597.

Marcus, G. F. (2001). The algebraic mind. Cambridge, MA: MIT Press.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. Cognitive Psychology, 21, 398-421.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. Cognitive Psychology, 23, 558-614.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. Cognitive Psychology, 23, 1-44.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1--86.

McClelland, J. L., & Patterson, K. (2002a). Rules or connections in past-tense inflections: What does the evidence rule out? Trends in Cognitive Sciences, 6, 465-472.

McClelland, J. L., & Patterson, K. (2002b). `Words or Rules' cannot exploit the regularity in exceptions. Trends in Cognitive Sciences, 6, 464-465.

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(5), 375-407.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. Psychological Review, 108(1), 113-148.

Pinker, S. (1991). Rules of language. Science, 253, 530-535.

Pinker, S., & Ullman, M. T. (2002a). The past and future of the past tense. Trends in Cognitive Sciences, 6, 456-463.

Pinker, S., & Ullman, M. T. (2002b). Combination and structure, not gradedness, is the issue. Trends in Cognitive Sciences, 6, 472-474.

Plaut, D. C., McClelland, J. L., & Seidenberg, M. S. (1995). Reading exception words and pseudowords: Are two routes really necessary? In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), Connectionist models of memory and language(p. 145-159). London: UCL Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. Psychological Review, 103, 56-115.

Plunkett, K., & Marchman, V. A. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. Cognition, 38, 43-102.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. Cognitive Psychology, 45(1), 45-94.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition(Vol. 2, p. 216-271). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations & Volume II: Psychological and biological models. Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Shirai, Y., & Anderson, R. W. (1995). The acquisition of tense-aspect morphology. a prototype account. Language, 71, 743-762.

Taatgen, N., & Anderson, J. A. (in press). Why do children learn to say "broke"? A model of learning the past tense without feedback. Cognition.

Ullman, M. T., Corkin, S., Coppola, M., Hicock, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). 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.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. Psychological Review, 108, 550-592.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proceedings of the National Academy of Science, 92, 930-933.

Acknowledgements