abstract algebra-like rules? *Trends Cognit. Sci.* 3, 166–168
7 Marcus, G.F. (1999) Connectionism: with or without rules? *Trends Cognit. Sci.* 3, 168–170
8 Altmann, G.T.M. and Dienes, Z. (1999) Rule learning by seven-month-old infants and neural networks *Science* 284, 875
9 Marcus, G.F. (1999) Rule learning by seven-month-old infants and neural networks. Response to Altmann and Dienes *Science* 284, 875
10 Jucszyk, P. (1997) *The Discovery of Spoken Language*, MIT Press
11 Christiansen, M.H. and Curtin, S.L. The power of statistical learning: no need for algebraic rules, in *Proc. 21st Annu. Conf. Cognit. Sci. Soc.* (Mahwah, NJ), Erlbaum (in press) (a Web-based version can be found at http://www.siu.edu/~psycho/faculty/morten/statlearn.html)
12 Christiansen, M.H. and Curtin, S.L. (1999) Transfer of learning: rule acquisition or statistical learning? *Trends Cognit. Sci.* 3, 289–290
13 Marcus, G.F. (1999) Reply to Christiansen and Curtin *Trends Cognit. Sci.* 3, 290–291
14 Pinker, S. and Prince, A. (1988) On language and connectionism: analysis of a parallel distributed processing model of language acquisition *Cognition* 28, 73–194

Reply to Seidenberg and Elman

We completely agree with Seidenberg and Elman[1] that what is important is 'understanding the behaviour of the network and the principles that govern it', but are mystified by their discussion of rules. Seidenberg and Elman's claim that 'the concept of "rule" has been altered to conform to the behaviour of connectionist networks' is a clever, ironic twist, but it actually has no substance, and reflects only Seidenberg and Elman's misunderstanding of our earlier point. As I emphasized in a series of recent articles, only a subset of neural networks implement rules[2–4]; not all of them do. As we stated before[2], the ones that depend on 'open-ended abstract relationships for which we can substitute arbitrary items' are the ones that implement rules. As it turns out, the *neural network* part of Seidenberg and Elman's system does not implement a rule. But, crucially, another part of the system that Seidenberg and Elman propose *does* implement a rule: the external teacher (this is what we meant by a 'hidden rule'). In particular, the external teacher incorporates a universally open-ended rule of the sort, *for all syllables x, y, if* $x = y$ *then output 1 else output 0*. This corresponds exactly to our characterization of rules. Since part of the system contains a rule, it follows that the system *as a whole* contains a rule.

Seidenberg and Elman also point to models of Altmann and Dienes[5] and Christiansen and Curtin[6]. We have discussed the Altmann and Dienes model elsewhere[7]. The pointer to the Christiansen and Curtin model is at best premature: the difference underlying their crucial result was small, it has not been shown to be statistically significant, and it has not yet been replicated (for a more detailed response see my reply to Christiansen and Curtin, in this issue[8]).

We do not expect this discussion to die down any time soon; clearly, advocates of certain types of connectionism believe that a lot is at stake. But the proliferation of alternative models should not distract us from noticing limits on particular classes of models where such limits do exist. For it is only by taking limits seriously that we can hope to build better models.

Gary F. Marcus
Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.
tel: +1 212 998 3551
fax: +1 212 995 4292
e-mail: gary.marcus@nyu.edu
http://www.psych.nyu.edu/~gary

References
1 Seidenberg, M.S. and Elman, J.L. (1999) Networks are not 'hidden rules' *Trends Cognit. Sci.* 3, 288–289
2 Marcus, G.F., Vijayan, S., Rao, S.B. and Vishton, P.M. (1999) Rule learning in 7-month-old infants *Science* 283, 77–80
3 Marcus, G.F. (1998) Rethinking eliminative connectionism *Cognit. Psychol.* 37, 243–282
4 Marcus, G.F. (1999) Do infants learn grammar with algebra or statistics? Response to Seidenberg & Elman, Negishi, and Eimas *Science* 284, 436–437
5 Altmann, G.T.M. and Dienes, Z (1999) Rule learning by seven-month-old infants and neural networks *Science* 284, 875
6 Christiansen, M.H. and Curtin, S.L. (1999) Transfer of learning: rule acquisition or statistical learning? *Trends Cognit. Sci.* 3, 289–290
7 Marcus, G.F. (1999) Rule learning by seven-month-old infants and neural networks. Response to Altmann and Dienes *Science* 284, 875
8 Marcus, G.F. (1999) Reply to Christiansen and Curtin *Trends Cognit. Sci.* 3, 290–291

Transfer of learning: rule acquisition or statistical learning?

Thirty years ago, Arthur Reber demonstrated that adults show evidence of transfer of learning in artificial-language experiments, in which the surface vocabulary is changed between training and test items[1]. In a series of experiments, Marcus and colleagues[2] demonstrated that infants as young as seven months old also show evidence of transfer of learning, but incorrectly conclude that the infants were extracting abstract algebraic rules rather than encoding statistical regularities. In contrast, a recent comprehensive review of the artificial-language-learning literature has demonstrated that transfer does not entail the involvement of abstract rules[3].

In Marcus *et al.*'s most persuasive demonstration of transfer of learning (Experiment 3 in Ref. 2), the infants were first trained on syllable sequences that followed either an AAB or ABB pattern (e.g. '*le-le-je*' versus '*le-je-je*'). The infants were then presented with sequences of novel syllables, either consistent or inconsistent with the training pattern. The infants showed a preference for the inconsistent items, thus demonstrating transfer between the different syllable vocabularies used in habituation and testing. Because there was no phonological overlap between training items and test items, Marcus *et al.* concluded that a statistical learning device could not account for these transfer results without implementing algebraic-like rules (see also the responses by Marcus[4,5] to commentaries by Seidenberg and Elman[6] and by McClelland and Plaut[7]). However, we suggest that statistical knowledge acquired in the service of learning to segment fluent speech into words might provide the basis for these transfer effects in much the same way as knowledge acquired in the process of learning to read can be used to perform experimental tasks such as lexical decision.

Using an existing simple recurrent-network[8] model of early infant speech segmentation[9] (Fig. 1), we tested this suggestion and successfully modeled the Marcus *et al.* results[10]. Importantly, no modifications were made to the original model, which learned to segment speech by integrating different kinds of probabilistic information derived from the speech stream (phonology, lexical stress and utterance-boundary information). Moreover, the simulation closely replicated the experimental conditions during both habituation and testing. The internal representations of the model were recorded at the end of each test item,

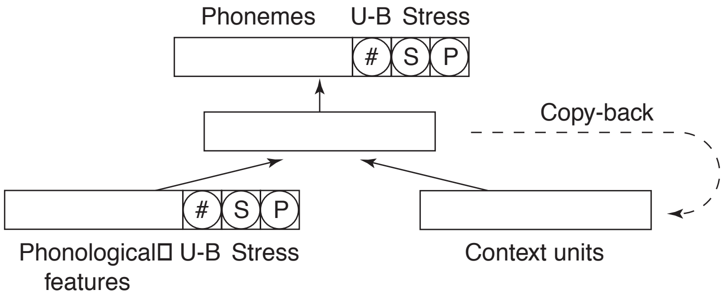

Fig. 1. An illustration of the simple recurrent network model of Christiansen *et al.*[9] used to model the Marcus *et al.* infant data[2]. Solid lines indicate trainable weights, whereas the dashed line denotes copy-back weights (which are always 1). The input to the model consisted of three kinds of information derived from a corpus of child-directed speech: (1) phonology represented in terms of 11 phonetic features on the input and 36 phonemes on the output; (2) utterance-boundary information represented as an extra feature marking utterance endings (U-B); and (3) lexical stress coded over two units as either no stress, secondary stress (S) or primary stress (P). The original network was trained to predict the next phoneme in a sequence as well as the appropriate values for the utterance boundary and stress units. (Modified from Ref. 9.)

and submitted to a two-group discriminant analysis. The results showed that these internal representations incorporated sufficient information to distinguish reliably between items that were either consistent or inconsistent with the habituation stimuli. Further analyses of the model's word-segmentation performance revealed that the model was better at segmenting out the words in the inconsistent items. This would make the inconsistent items more salient and therefore explain why the infants preferred these to the consistent items. Thus the transfer effects that Marcus *et al.* report can be readily accounted for by assuming that the infants' behavioral responses are based on statistical learning, similar to the above connectionist model.

All too often statistical-learning approaches – including connectionist models – are forced into a behavioristic mold[11]: only input–output relations are said to matter. However, the proponents of connectionist-style statistical learning have also taken part in the cognitive revolution and therefore posit internal representations mediating between input and output. Indeed, the internal representations of the above model provided a crucial source of information for the modeling of the infants' behavior in the Marcus *et al.* study. Another oversight of the critics of connectionism relates to the importance of integrating multiple sources of information within a single statistical-learning device[9,12]. It was this kind of information integration that enabled the above model to explain the infants' preference for the inconsistent items because its performance did not rely only on phonological information. Thus, a more sophisticated approach to statistical learning and connectionist modeling is needed to reveal their true power. Once such an approach is adopted it becomes clear that the impressive learning abilities of the infants in the Marcus *et al.* study do not require the postulation of abstract algebraic rules.

Morten H. Christiansen
Department of Psychology, Southern Illinois University, Carbondale, IL 62901-6502, USA.
tel: +1 618 453 3547
fax: +1 618 453 3563
e-mail: morten@siu.edu

Suzanne Curtin
Department of Linguistics, University of Southern California, Los Angeles, CA 90089-1693, USA.
e-mail: curtin@gizmo.usc.edu

References

1 Reber, A.S. (1969) Transfer of syntactic structure in synthetic languages *J. Exp. Psychol.* 81, 115–119
2 Marcus, G.F. *et al.* (1999) Rule learning in seven-month-old infants *Science* 283, 77–80
3 Redington, M. and Chater, N. (1996) Transfer in artificial grammar learning: a re-evaluation *J. Exp. Psychol. Gen.* 125, 123–138
4 Marcus, G.F. (1999) Do infants learn grammar with algebra or statistics? Response to Seidenberg & Elman, Negishi, and Eimas *Science* 284, 436–437
5 Marcus, G.F. (1999) Connectionism: with or without rules? Response to J.L. McClelland and D.C. Plaut (1999) *Trends Cognit. Sci.* 3, 168–170
6 Seidenberg, M.S. and Elman, J.L. (1999) Do infants learn grammar with algebra or statistics? Letter *Science* 284, 434–435
7 McClelland, J.L. and Plaut, D.C. (1999) Does generalization in infant learning implicate abstract algebra-like rules? *Trends Cognit. Sci.* 3, 166–168
8 Elman, J.L. (1990) Finding structure in time *Cognit. Sci.* 14, 179–211
9 Christiansen, M.H., Allen, J. and Seidenberg, M.S. (1998) Learning to segment using multiple cues: a connectionist model *Lang. Cognit. Processes* 13, 221–268
10 Christiansen, M.H. and Curtin, S.L. The power of statistical learning: no need for algebraic rules *Proc. 21st Annu. Conf. Cognit. Sci. Soc.* (Mahwah, NJ), Erlbaum (in press) (a Web-based version can be found at http://www.siu.edu/~psycho/faculty/morten/statlearn.html)
11 Pinker, S. (1999) Out of the minds of babes *Science* 283, 40–41
12 Redington, M. and Chater, N. (1997) Probabilistic and distributional approaches to language acquisition *Trends Cognit. Sci.* 1, 273–281

Reply to Christiansen and Curtin

Christiansen and Curtin (in this issue[1], described in more detail in[2]) offer two results that they take to be arguments against our arguments[3–9] about simple recurrent networks. The first result concerns a discriminant analysis which shows that the 'internal representations [in a simple recurrent network] incorporated sufficient information to reliably distinguish between items that were either consistent or inconsistent with the habituation stimuli'. In effect, what Christiansen and Curtin did was to plot twelve sentences with either AAB or ABB syllable sequences in an 80-dimensional space, and then ask whether it was possible to draw a line anywhere in those 80 dimensions that would divide the AAB test items from the ABB test items.

It is likely that in many dimensions, the AAB items and the ABB items overlap with one another; wherever there is overlap, one could not, by definition, separate the AAB items from the AAB items. What Christiansen and Curtin's first result shows us is that not all the dimensions are like that; instead, there exists at least one dimension in which the AAB and ABB items do not overlap; in that dimension (or set of dimensions), one could draw a line between the AAB and ABB items.

While we do not at all doubt that this is the case, the result is all but a statistical certainty – excluding repetitions of identical sentences, there are only four distinct data points ('*ba-ba-po*' and '*ko-ko-ga*' versus '*ba-po-po*' and '*ko-ga-ga*') to be separated, and 80 chances (one per hidden unit) to draw such a line. So the result might well have happened even by chance.

To protect against this possibility, Christiansen and Curtin ran a control in which they randomly reassigned three vectors from Group A to Group B, finding that one could no longer draw

1364-6613/99/$ – see front matter © 1999 Elsevier Science. All rights reserved.

a line that perfectly splits the two groups. But this control stacked the deck in Christiansen and Curtin's favor by asking the discriminant analysis to do the impossible: to succeed, it would have had to separate two instances of some sentence, say '*ba-ba-po*' from the third instance of the very same sentence. As a consequence, the control confounds repetition of sentences with categorization of sentences, and hence shows nothing about whether the initial result is meaningful. (A more appropriate control to assess the meaningfulness of the discriminant analysis might have been to ask whether a line could be drawn between the six sentences starting with the word *ba* and the six sentences starting with the word *ko*.)

Even if the discriminant analysis were real, what one would really want to do is to show that the internal representations play a causal role in behavior; Christiansen and Curtin's second analysis, which attempts to do this, is therefore the crucial one. In particular, in this analysis they showed that if our materials are presented to a simple recurrent network as a single uninterrupted string with no pauses between words, the model was slightly better (80% versus 75%) at 'segmenting out the words in the inconsistent items'. From this, they argued that, 'This would make the inconsistent items more salient and therefore explain why the infants preferred these to the consistent items'[1].

But even if we accepted the premise that infants would look longer at items that were easier to segment, Christiansen and Curtin's position runs into at least two serious problems. First, their result, if it is real, might rely on setting up our task as a segmentation problem in which there are no pauses between words – but, our materials included a 250-ms pause between words (see footnote 11, Ref. 3). Christiansen and Curtin's attempt to explain the infants' performance as a segmentation task thus does not fit the actual task, and we suspect that their effect – if it is real – would go away if these gaps between words were included.

Second, an even more pressing issue is to see whether the result is real. Given that in the experimental stimuli grammatical consistency was not correlated with any sort of segmentation cue, it might turn out that Christiansen and Curtin's result is simply a chance occurrence that has been wrongly interpreted as meaningful: the differences observed were small, statistically significance has not yet been reported, and the simulations have not yet been replicated. In the absence of assurances that their behavioural result is robust, it does not merit further discussion at present.

Gary F. Marcus
Department of Psychology,
New York University,
6 Washington Place, New York,
NY 10003, USA.
tel: +1 212 998 3551
fax: +1 212 995 4292
e-mail: gary.marcus@nyu.edu
http://www.psych.nyu.edu/~gary

Acknowledgement

I thank Morten Christiansen, Bob Hadley, Pat Shrout and Zsofia Zvolenszky for helpful discussion.

References

1 Christiansen, M.H. and Curtin, S.L. (1999) Transfer of learning: rule acquisition or statistical learning? *Trends Cognit. Sci.* 3, 289–290

2 Christiansen, M.H. and Curtin, S.L. The power of statistical learning: no need for algebraic rules *Proc. 21st Annu. Conf. Cognit. Sci. Soc.* (Mahwah, NJ), Erlbaum (in press) (a Web-based version can be found at http://www.siu.edu/~psycho/faculty/morten/statlearn.html)

3 Marcus, G.F., Vijayan, S., Rao, S.B. and Vishton, P.M. (1999) Rule learning in seven-month-old infants *Science* 283, 77–80

4 Marcus, G.F. (1998) Rethinking eliminative connectionism *Cognit. Psychol.* 37, 243–282

5 Marcus, G.F. (1998) Can connectionism save constructivism? *Cognition* 66, 153–182

6 Marcus, G.F. (1997) *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations* (Review) *Trends Cognit. Sci.* 1, 318–319

7 Marcus, G.F. (1999) Do infants learn grammar with algebra or statistics? Response to Seidenberg & Elman, Negishi, and Eimas *Science* 284, 436–437

8 Marcus, G.F. (1999) Connectionism: with or without rules? Response to J.L. McClelland and D.C. Plaut *Trends Cognit. Sci.* 3, 168–170

9 Marcus, G.F. (1999) Rule learning by seven-month-old infants and neural networks. Response to Altmann and Dienes *Science* 284, 875

Seeing (and hearing) the wood for the trees

Third International Conference on Cognitive and Neural Systems, 26–29 May 1999, Boston University, MA, USA.

One of the central issues in the study of perception has been how we should conceptualize the relationship between the parts and the whole. Since the Gestalt psychologists first raised this problem it has evolved into several distinct issues. These include the relative priority of features or global structure, the relative importance of bottom-up or top-down processing, and the importance of modularity or the binding of different features together during processing. These issues were continually revisited throughout the presentations at this conference.

For example, in the domain of speech perception, Steven Greenberg (Berkeley, CA, USA) proposed a syllable-centric perspective on spoken language[1]. He argued that syllables, although very difficult to define, offer a better basis than phonemes on which to develop a semantic network. Greenberg presented empirical data indicating that analysis windows of syllabic length (150–300 ms) are important for speech perception, and that fine spectral information might not be required for decoding acoustic signals. The emphasis on the syllable/context level and redundancy in spectral channels is reminiscent of the latest trend in visual perception, that is, the growing realization of the significance of intermediate-level surface representation as opposed to the earlier level of local feature detectors.

Stephen Grossberg (Boston, MA, USA) echoed this theme on a larger scale[2]. Grossberg characterized the critical issues of phonetic restoration (the observed fact that a missing phoneme in a familiar word is perceptually restored so that the word sounds complete) as follows: How does the future influence the past? And how does meaning affect the perception of phonemes? According to Grossberg, top-down expectation is required to resolve these questions, athough the question then arises as to how we can realistically implement this top-down influence in intelligent artificial systems. Here Grossberg appealed to the development of resonant states between

 © 1999 Elsevier Science. All rights reserved. PII: S1364-6613(99)01359-5