

# Rule Learning by Seven-Month-Old Infants and Neural Networks

Zoltán Dienes

Laboratory of Experimental Psychology,  
University of Sussex,  
Brighton BN1 9QG, United Kingdom

Gary F. Marcus *et al.* (1) familiarized 7-month-old infants with sequences of syllables generated by an artificial grammar; the infants were then able to discriminate between sequences generated both by that grammar and another, even though sequences in the familiarization and test phases employed different syllables. Marcus *et al.* stated that their infants were representing, extracting, and generalizing abstract algebraic rules. This conclusion was motivated also by their statement that the infants' discrimination could not be performed by a popular class of simple neural network model.

Marcus *et al.* make a number of statements regarding the supposed inability of statistical learning mechanisms, including neural networks, to account for their data. One model they describe (2) was developed, however, to model precisely the kinds of abstract generalizations exhibited by the infants in the report. Marcus *et al.* state that this model cannot account for their data because, unlike the infants, it relies on being supplied with attested examples of sentences that are acceptable in the artificial language used during the test phase. This is not the case.

Our model used a simple recurrent network (3) with an extra encoding layer between the input and hidden layers of nodes. During training, the network was presented with grammatical sequences of syllables (each input node corresponded to a particular syllable). The network's task was to predict the next syllable. Weights on connections between the nodes were adjusted with the use of back-propagation. At test, the weights on connections to the hidden layer were frozen (simulating an adaptive learning procedure). Both grammatical and ungrammatical sequences of hitherto unseen syllables were then presented to the input layer, using input nodes that had not been used for training. The network's classification of these sequences [determined by an equivalent of the Luce choice rule (4)] did not differ from human participants' above-chance classification of the same stimuli. This was achieved, contrary to the description given by Marcus *et al.*, without pre-test exposure to any test sequences, without feedback at test on which sequences were grammatical and which ungrammatical, and when the input nodes corresponded to individual words in the language (contrary to a further statement in their report

concerning limitations on generalization).

Subsequent simulations using the same network (5) demonstrated that above-chance discrimination at test between "grammatical" and "ungrammatical" sequences can be achieved (within certain parameters) even when the only basis for discrimination is the difference in "repetition structure" (for example, the ABA and ABB manipulation in the report). The statement by Marcus *et al.* that the ability to represent repetition patterns such as ABB or AAB is outside the scope of most neural network models of language is unlikely to be correct in light of our findings and demonstrations of recurrent networks' ability to learn context-free grammars generating AB, AABB, AAABBB . . . (6).

We have also simulated the findings in the report (1) directly, using their own stimuli. We trained eight versions of our network on the ABA grammar and eight on the ABB grammar. At test, each network was presented with ABA and ABB test sentences in random order, using input nodes that had not been used in training (thereby simulating a change in vocabulary). As each network "saw" each successive test sequence, we correlated the network's prediction of what it would see next with the next input, and calculated also the Euclidian distance between the two. With learning rate and momentum set at 0.5 and 0.01, respectively, and 10 iterations around each test item, we found significantly higher correlations for congruent sequences than for incongruent ones [ $F(1,15) = 20.8$ ,  $P = 0.0004$ ], and a significantly smaller euclidian distance between prediction and target for congruent targets than for incongruent ones [ $F(1,15) = 23.1$ ,  $P = 0.0002$ ]. Like the infants studied by Marcus *et al.*, our networks successfully discriminated between the test stimuli.

The conclusions by Marcus *et al.* stated in the report are premature; a popular class of neural network can model aspects of their own data, as well as substantially more complex data than those in the report. The cognitive processes of 7-month-old infants may not be so different from statistical learning mechanisms after all.

Gerry T. M. Altmann

Department of Psychology,  
University of York Heslington,  
York YO10 5DD, United Kingdom  
E-mail: g.altmann@psych.york.ac.uk

## References and Notes

1. G. F. Marcus, S. Vijayan, S. Bandi Rao, P. M. Vishton, *Science* **283**, 77 (1999).
2. Z. Dienes, G. T. M. Altmann, S.-J. Gao, in *Neural Computation and Psychology*, L. S. Smith and P. J. B. Hancock, Eds. (Springer-Verlag, New York, 1995); G. T. M. Altmann, Z. Dienes, A. Goode, *Exp. Psychol. Learn. Mem. Cognit.* **21**, 899 (1995).
3. J. L. Elman, *Cognit. Sci.* **14**, 179 (1990).
4. R. D. Luce, *Individual Choice Behaviour* (Wiley, New York, 1959).
5. Z. Dienes and G. T. M. Altmann, S.-J. Gao, *Cognit. Sci.*, in press.
6. P. Rodriguez, *Center for Research in Language Newsletter, University of California*, No. 10 (1995).

3 March 1999; accepted 13 April 1999

*Response:* Although I concede that in our report (1) we misunderstood the model of Dienes, Altmann, and Gao, there are still three good reasons to doubt that their model provides the correct account.

First, whether the model can be said to capture our data depends on how one interprets the outputs of the model. If one interprets the model in terms of what word is most active at any given moment (a standard way of analyzing models of this general class), one finds that the model does not "learn" the training grammar, but instead oscillates between (say) the ABA and ABB grammars. In contrast, if one uses the method of Altmann and Dienes for interpreting the model, one finds that the model successfully discriminates between the consistent and inconsistent items. Thus, the model could have been made compatible with our results (by interpreting the outputs as do Altmann and Dienes) or with a result in which the infants could not make the discrimination (by interpreting the model's output to be the most active unit).

Even putting aside this issue of how to interpret the model, one is still left with a purpose-built model that makes a number of nonstandard assumptions that limit its generality and plausibility. The "success" of the model depends on iterating each test sentence multiple times before moving on to the next sentence (a technique not generally used in connectionist simulations of language and cognition), and an external mechanism that can freeze a specific subset of connection weights. (In other words, this external device must "turn off" learning in one part of the model, but not another.) It is unclear what sort of neural system could implement this in the brief period of time which the infants have in our experiments, and unclear whether the architecture proposed by Altmann and Dienes could be used in other cognitive tasks.

Finally, and most important, my preliminary work with the model of Altmann and Dienes suggests that what the model does is

## TECHNICAL COMMENT

not to genuinely abstract a rule, but rather to map the encodings of one set of words onto the encodings of another set of words. To tease apart these two possibilities, I conducted a slightly different experiment, in which the model was habituated on the same 16 ABA sentences as our infants in Experiment 1 in our report (“ga ti ga,” and so on), but tested on a slightly different set of test trials. In the first two test trials, the model was tested on the item “wo fe wo”; after a number of iterations, the model mapped “wo” onto (say) “ga” and “fe” onto (say) “ti.” Next, I trained the model on either “fe wo wo” or “fe wo fe.” With the use of the output interpretation technique developed by Altmann and Dienes, I found that the model would “look longer” at “fe wo fe” than at “fe wo wo,”

presumably because the model is strongly driven by information about “wo” appearing as the third word in the sentence. An interesting, open question is whether infants, too, would do this, or whether they would do the opposite, looking longer at “fe wo wo” than at “fe wo fe.”

**Gary F. Marcus**

*Department of Psychology,  
New York University,  
New York, NY 10003, USA  
E-mail: gary.marcus@nyu.edu*

### References

1. G. F. Marcus, S. Vijayan, S. Bandi Rao, P. M. Vishton, *Science* **283**, 77 (1999).

29 March 1999; accepted 13 April 1999