

Rule Learning by Seven-Month-Old Infants

G. F. Marcus,^{*} S. Vijayan, S. Bandi Rao, P. M. Vishton

A fundamental task of language acquisition is to extract abstract algebraic rules. Three experiments show that 7-month-old infants attend longer to sentences with unfamiliar structures than to sentences with familiar structures. The design of the artificial language task used in these experiments ensured that this discrimination could not be performed by counting, by a system that is sensitive only to transitional probabilities, or by a popular class of simple neural network models. Instead, these results suggest that infants can represent, extract, and generalize abstract algebraic rules.

What learning mechanisms are available to infants on the cusp of language learning? One learning mechanism that young infants can exploit is statistical in nature. For example, Saffran *et al.* (1) found that the looking behaviors of 8-month-old infants indicated a sensitivity to statistical information inherent in sequences of speech sounds produced in an artificial language—for example, transitional probabilities, which are estimates of how likely one item is to follow another. In the corpus of sentences "The boy loves apples. The boy loves oranges." the transitional probability between the words "the" and "boy" is 1.0 but the transitional probability between the words "loves" and "apples" is $1/2 = 0.5$.

It has been suggested that mechanisms that track statistical information, or connectionist models that rely on similar sorts of information [for example, the simple recurrent network (SRN) (2)], may suffice for language learning (3). The alternative possibility considered here is that children might possess at least two learning mechanisms, one for learning statistical information and another for learning "algebraic" rules (4)—open-ended abstract relationships for which we can substitute arbitrary items. For instance, we can substitute any value of x into the equation $y = x + 2$. Similarly, if we know that in English a sentence can be formed by concatenating any plural noun phrase with any verb phrase with plural agreement, then

as soon as we discover that "the three blickets" is a well-formed plural noun phrase and that "reminded Sam of Tibetan art" is a well-formed verb phrase with plural agreement, we can infer that "The three blickets reminded Sam of Tibetan art." is a well-formed sentence.

To date, however, there has been no direct empirical test for determining whether young infants can actually learn simplified versions of such algebraic rules. A number of previous experiments drawn from the literature of speech perception (not aimed at the question of rule learning) are consistent with the possibility that infants might learn algebraic rules, but each of these prior experiments could be accounted for by a system that extracted only statistical tendencies. For example, infants who are habituated to a series of two-syllable words attend longer when confronted with a three-syllable word (5). An infant who attended longer to a three-syllable word might have noticed a violation of a rule (for example, "all the words here are two syllables"), but an infant could also have succeeded with a statistical device that noted that the three-syllable word had more syllables than the average number of syllables in the preceding utterance. Similarly, Gomez and Gerken (6) found that infants who were habituated to a set of sentences constructed from an artificial grammar (VOT-PEL-JIC; PEL-TAM-PEL-JIC) could distinguish between new sentences that were consistent with this grammar (VOT-PEL-TAM-PEL-JIC) from new sentences that were not consistent (VOT-TAM-PEL-RUD-JIC). Such learning might reflect the acquisition of rules, but because all the test sentences were constructed with the same words as in the habituation sentences (albeit rearranged), in these test sentences it was possible to distinguish the test sentence on the basis of

statistical information such as transitional probabilities (for example, in the training corpus, VOT was never followed by TAM)—without recourse to a rule.

We tested infants in three experiments in which simple statistical or counting mechanisms would not suffice to learn the rule that was generating the sequences of words. In each experiment, infants were habituated to three-word sentences constructed from an artificial language (7) and then tested on three-word sentences composed entirely of artificial words that did not appear in the habituation. The test sentences varied as to whether they were consistent or inconsistent with the grammar of the habituation sentences. Because none of the test words appeared in the habituation phase, infants could not distinguish the test sentences based on transitional probabilities, and because the test sentences were the same length and were generated by a computer, the infant could not distinguish them based on statistical properties such as number of syllables or prosody.

We tested infants with the familiarization preference procedure as adapted by Saffran *et al.* (1, 8, 9); if infants can abstract the underlying structure and generalize it to novel words, they should attend longer during presentation of the inconsistent items than during presentation of consistent items.

Subjects were 7-month-old infants, who were younger than those studied by Saffran *et al.* but still old enough to be able to distinguish words in a fluent stream of speech (8). In the first experiment, 16 infants were randomly assigned to either an "ABA" condition or an "ABB" condition. In the ABA condition, infants were familiarized with a 2-min speech sample (10) containing three repetitions of each of 16 three-word sentences that followed an ABA grammar, such as "ga ti ga" and "li na li." In condition ABB, infants were familiarized with a comparable speech sample in which all training sentences followed an ABB grammar, such as "ga ti ti" and "li na na" (11).

In the test phase, we presented infants with 12 sentences that consisted entirely of new words, such as "wo fe wo" or "wo fe fe" (12). Half the test trials were "consistent sentences," constructed from the same grammar as the one with which the infant was familiarized (an ABA test sentence for infants trained in the ABA condition and an ABB sentence for infants trained in the ABB condition), and half the test trials were "inconsistent sentences" that were constructed from the grammar on which the infant was not trained (13).

We found that 15 of 16 infants showed a preference for the inconsistent sentences

G. F. Marcus, S. Vijayan, S. Bandi Rao, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA. P. M. Vishton, Department of Psychology, Amherst College, Amherst, MA 01002, USA.

^{*} To whom correspondence should be addressed. E-mail: gary.marcus@nyu.edu.

(14), which was indicated by their looking longer at the flashing side light during presentations of those sentences (15) (Table 1).

Although each of the test words in experiment 1 was new, the sequence of phonetic features in the test overlapped to some extent with the sequence of phonetic features in the habituation items. For example, in the ABA condition three habituation sentences contained a word starting with a voiced consonant followed by a word starting with an unvoiced consonant. Each of these three sequences ended with a word that contained a voiced consonant. An infant who was thus expecting the sequence voiced-unvoiced-voiced would be surprised by the inconsistent tests items (each of which was voiced-unvoiced-unvoiced) but not by the consistent items (each of which was voiced-unvoiced-voiced). To rule out the possibility that infants might rely on learning sequences of particular phonetic features rather than deriving a more abstract rule, we conducted a second experiment with the same grammars as in the first experiment but with a more carefully constructed set of words. In experiment 2, then, the set of phonetic features that distinguished the test words from each other did not distinguish the words that appeared in the habituation sentences (16). For example, the test words varied in the feature of voicing (for example, if the "A" word was +voiced, the "B" word was -voiced), whereas the habituation words did not vary on the feature of voicing (they were all +voiced). Thus, the habituation items provided no direct information about the relationship between voiced and unvoiced consonants; the same holds for each of the phonetic features that varied in the test items. As in experiment 1, 15 of 16 infants looked longer during the presentation of the inconsistent items than during the presentation of the consistent items (17) (Table 1).

Rather than encoding the entire ABA or ABB rule, the infants could have habituated to a single property that distinguishes these grammars. Strings from the ABB grammar contain immediately reduplicated elements (for example, "ti ti"), whereas strings from the ABA grammar do not. In a third experiment, we compared sentences constructed from the ABB grammar with sentences constructed from an AAB grammar (18, 19); because reduplication was contained in both grammars, the infants could not distinguish these grammars solely on the basis of information about reduplication (20). As in the first two experiments, infants (this time, 16 of 16) looked longer during presentation of the inconsistent items than during presentation of the consistent items (21) (Table 1).

Our results do not call into question the existence of statistical learning mechanisms but show that such mechanisms do not exhaust the child's repertoire of learning mechanisms. A system that was sensitive only to transitional probabilities between words could not account for any of these results, because all the words in the test sentences are novel and, hence, their transitional probabilities (with respect to the familiarization corpus) are all zero. Similarly, a system that noted discrepancies with stored sequences of words could not account for the results in any of the three experiments, because both the consistent items and the inconsistent items differ from any stored sequences of words. A system that noted discrepancies with stored sequences of phonetic features could account for the results in experiment 1 but not those in experiments 2 and 3. A system that could count the number of reduplicated elements and notice sentences that differ in the number of reduplicated elements could account for the results in experiments 1 and 2, but it could not account for infants' performance in experiment 3.

Likewise, we found in a series of

simulations that the SRN is unable to distinguish the inconsistent and consistent sentences, because the network, which represents knowledge in terms of a set of connection weights, learns by altering network connection weights for each word independently (22). As a result, there is no generalization to novel words. Such networks can simulate knowledge of grammatical rules only by being trained on all items to which they apply; consequently, such mechanisms cannot account for how humans generalize rules to new items that do not overlap with the items that appeared in training (23, 24).

We propose that a system that could account for our results is one in which infants extract abstract algebra-like rules that represent relationships between placeholders (variables), such as "the first item X is the same as the third item Y," or more generally, that "item I is the same as item J." In addition to having the capacity to represent such rules, our results appear to show that infants have the ability to extract those rules rapidly from small amounts of input and to generalize those rules to novel instances. If our position is correct, then infants possess at least two distinct tools for learning about the world and attacking the problem of learning language: one device that tracks statistical relationships such as transitional probabilities and another that manipulates variables, allowing children to learn rules. Even taken together, these tools are unlikely to be sufficient for learning language, but both may be necessary prerequisites.

REFERENCES AND NOTES

1. J. Saffran, R. Aslin, E. Newport, *Science* **274**, 1926 (1996).
2. J. L. Elman, *Cogn. Sci.* **14**, 179 (1990).
3. E. Bates and J. Elman, *Science* **274**, 1849 (1996); M. S. Seidenberg, *ibid.* **275**, 1599 (1997).
4. N. A. Chomsky, *Rules and Representations* (Columbia Univ. Press, New York, 1980); S. Pinker and A. Prince, *Cognition* **28**, 73 (1988); S. Pinker, *Science* **253**, 530 (1991); G. F. Marcus, U. Brinkmann, H. Clahsen, R. Wiese, S. Pinker, *Cogn. Psychol.* **29**, 186 (1995) . G. F. Marcus, *The Algebraic Mind* (MIT Press, Cambridge, MA, 1999), in press.
5. R. Bijeljac-Babic, J. Bertoncini, J. Mehler, *Dev. Psychol.* **29**, 711 (1993).
6. R. L. Gomez and L.-A. Gerken, in *Boston University Conference on Language Development 21*, E. Hughes, M. Hughes, A. Greenhill, Eds. (Cascadia Press, Somerville, MA, 1997), p. 194.
7. We leave open the question of whether infants interpreted our materials as genuinely linguistic and thus also leave open the question of whether the mechanisms that acquire abstract rules are specific to language learning or are more generally used in many domains.

Table 1. Mean time spent looking in the direction of the consistent and inconsistent stimuli in each condition for experiments 1, 2, and 3, and significance tests comparing the listening times. Mean ages of the infants tested were 6 months 27 days (median, 6 months 24 days) in experiment 1, 7 months 1 day (median, 7 months) in experiment 2, and 7 months (median, 7 months 2 days) in experiment 3.

Exp.	Mean listening time (s) (SE)		Repeated measures analysis of variance
	Consistent sentences	Inconsistent sentences	
1	6.3 (0.65)	9.0 (0.54)	$F(14) = 25.7, P < 0.001$
2	5.6 (0.47)	7.35 (0.68)	$F(14) = 25.6, P < 0.005$
3	6.4 (0.38)	8.5 (0.5)	$F(14) = 40.3, P < 0.001$

8. P. Jusczyk and R. N. Aslin, *Cogn. Psychol.* **29**, 1 (1995).
9. Infants sat in a three-sided booth on the laps of their parents (parents wore headphones playing classical music so that they could not hear the stimulus materials) and listened to sounds generated off-line by a speech synthesizer. The booth had a yellow bulb on the center panel; each side panel had a red bulb. A speaker was behind each of the red bulbs. The speakers were connected to a G3 Power Macintosh computer that presented the stimuli and controlled the lights. During the familiarization phase, the yellow light flashed to draw the infant's attention to the center panel of the testing booth while the familiarization speech segment played from both speakers. After the familiarization ended, the infant was presented with test trials. At the beginning of each test trial, the central light was flashed. Once an observer (who also wore headphones playing music to mask the stimuli) indicated that the infant had fixated on the flashing light, the central light was turned off and one of the two side lights began flashing. When the observer indicated that the infant had turned toward the side light, the computer played a three-word test sentence from the speaker that was hidden behind the light, which repeated the test sentence over and over (with a 1.2- to 1.5-s pause between presentations of the test sentence) until either the infant had turned away for two continuous seconds or until 15 s had elapsed. The dependent measure was the total time that the infant spent looking at the light associated with the speaker. Infants who became fussy prior to completion of at least four test trials were not included in the statistical analyses.
10. The first six subjects (three in each condition) were familiarized with 3-min speech samples.
11. The 16 sentences that followed an ABA pattern were "ga ti ga," "ga na ga," "ga gi ga," "ga la ga," "li na li," "li ti li," "li gi li," "li la li," "ni gi ni," "ni ti ni," "ni na ni," "ni la ni," "ta la ta," "ta ti ta," "ta na ta," and "ta gi ta." The 16 sentences that followed the pattern ABB were "ga ti ti," "ga na na," "ga gi gi," "ga la la," "li na na," "li ti ti," "li gi gi," "li la la," "ni gi gi," "ni ti ti," "ni na na," "ni la la," "ta la la," "ta ti ti," "ta na na," and "ta gi gi." Vocalizations of the words used in the above sentences were created with a speech synthesizer, which is available at www.bell-labs.com/project/tts/voices-java.html. The vocalizations were then combined to form the sentences listed above by using a sound editor. A 250-ms pause was placed between consecutive words in each sentence. The sentences were presented in random order and separated by pauses of 1 s.
12. The 12 test trials, which were randomly ordered, included three repetitions of each of four test sentences, two following the ABB pattern ("wo fe fe" and "de ko ko") and two following the ABA pattern ("wo fe wo" and "de ko de").
13. Similar stimuli were used in a study of children's memory and attention [J. V. Goodstitt, P. A. Morse, J. N. Ver Hoeve, *Child Dev.* **55**, 903 (1984)]. That study does not, however, answer our question about rules, because it tested only how well an infant could remember target B in the context of sequences ABA versus AAB versus ABC and not whether infants familiarized with one of those sequences could distinguish it from another.
14. Results for the ABA and ABB conditions were combined, because there was no significant interaction between them, $F(1,14) = 0.15$.
15. Similar results involving transfer from one finite state grammar to another with the same structure but different words have been reported for adult subjects [A. Reber, *J. Exp. Psychol.* **81**, 115 (1969)] and for 11-month-old infants (R. L. Gomez and L.-A. Gerken, paper presented at the Annual Meeting of the Psychonomics Society, Philadelphia, PA, November 1997). These researchers, whose focus was not on rule learning, did not include the phonetic control we introduce in experiments 2 and 3.
16. The 16 habituation sentences that followed the ABA pattern were "le di le," "le je le," "le li le," "le we le," "wi di wi," "wi je wi," "wi li wi," "wi we wi," "ji di ji," "ji je ji," "ji li ji," "ji we ji," "de di de," "de je de," "de li de," "de we de"; ABB items were constructed with the same vocabulary. The test trials were "ba po ba," "ko ga ko" (consistent with ABA), "ba po po," and "ko ga ga" (consistent with ABB).
17. Results for the ABA and ABB conditions were combined, because there was no significant interaction between them, $F(1,14) = 1.95$.
18. In principle, an infant who paid attention only to the final two syllables of each sentence could distinguish the AAB grammar from the ABB grammar purely on the basis of reduplication, but they could not have succeeded in the experiment of Saffran *et al.* (1).
19. We thank an anonymous reviewer for suggesting this comparison. The vocabulary used to construct the test and familiarization items was the same as in experiment 2; hence, as in experiment 2, the phonetic features that distinguished the test words from each other did not vary in the habituation items.
20. The ability to extend reduplication to novel words appears to depend on an algebraic rule. To recognize that an item is reduplicated, a system must have the ability to store the first element and compare the second element to the first; the storage, retrieval, and inferential mechanisms that are involved may appear simple but are outside the scope of most neural network models of language and cognition. Conversely, adults are strongly sensitive to the presence of reduplication and its location in phonological constituents [I. Berent and J. Shimron, *Cognition* **64**, 39 (1997)]. For further discussion, see references cited in (22).
21. Results for the AAB and ABB conditions were combined, because there was no significant interaction between them, $F(1,14) = 0.002$.
22. The sort of generalizations that such models can draw are dictated by the choice of input representations. If input nodes correspond to words, the model cannot generalize the abstract pattern to new words; if the input nodes correspond to phonetic features, the model cannot generalize to words containing new phonetic features [G. F. Marcus, *Cognition* **66**, 153 (1998); *Cogn. Psychol.*, in press]. An appropriately configured SRN that represented each word by a set of nodes for phonetic features, if it were trained that a voiced consonant followed by an unvoiced consonant was always followed by a voiced consonant, could use memorized sequences of features as a basis to distinguish the test items in experiment 1. However, such a model could not account for the results of experiments 2 and 3, because in those experiments the feature sequences that the network learned about in the familiarization phase would not distinguish the test items.
23. An enhanced version of the SRN [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)] aims to model how speakers who are trained on one artificial language are able to learn a second artificial language that has the same structure more rapidly than a second artificial language that has a different structure. This model would not be able to account for our data, however, because the model relies on being supplied with attested examples of sentences that are acceptable in the second artificial language, whereas our infants succeeded in the absence of such information.
24. The problem is not with neural networks per se but with the kinds of network architectures that are currently popular. These networks eschew explicit representations of variables and relationships between variables; in contrast, some less widely discussed neural networks with a very different architecture do incorporate such machinery and thus might form the basis for learning mechanisms that could account for our data [J. E. Hummel and K. J. Holyoak, *Psychol. Rev.* **104**, 427 (1997)]. Our goal is not to deny the importance of neural networks but rather to try to characterize what properties the right sort of neural network architecture must have.
25. Supported in part by an Amherst College Faculty Research Grant to P.M.V. We thank L. Bonatti, M. Brent, S. Carey, J. Dalalakis, P. Gordon, B. Partee, V. Valian, and Z. Zvolenszky for helpful discussion, and P. Marcus and F. Scherer for their assistance in construction of the test apparatus. We also thank Bell Labs for making available to the public the speech synthesizer that we used to create our stimuli. Some subjects in experiment 1 were tested at Amherst College; all other

subjects were tested at New York University. The parents of all participants gave informed consent.

11 September 1998; accepted 16 November 1998

Out of the Minds of Babes

Steven Pinker*

It is a cliché of neuroscience that the brain works differently from a digital computer. But the report by Marcus *et al.* in this issue on page 77 (1) demonstrating "rule learning by seven-month-old infants" suggests that one of the mechanisms that makes computers intelligent--manipulating symbols according to rules--may be a basic mechanism of the human brain as well.

Hundreds of years before anyone knew anything about brains or computers, two very different conceptions arose of how the mind works:

"When a man reasons, he does nothing else but conceive a sum total from addition of parcels, or conceive a remainder from subtraction of one sum from another; which, if it be done by words, is conceiving of the consequence of the names of all the parts to the name of the whole, or from the names of the whole and one part to the name of the other part. ... For REASON is nothing but *reckoning*."

In this passage from *Leviathan*, written in 1651 (2), Thomas Hobbes uses "reckoning" in the original sense of "calculating" or "computing." For example, if the definition of "man" is "rational animal," and we are told that something is "rational" and an "animal" (names of parts), we can deduce it is a "man" (name of whole). If these symbols are represented as patterns of activity in the brain, and if some patterns trigger other patterns because of the way the brain is organized, then we have a theory of intelligence. That theory became the basis of the rationalist philosophy of Descartes and Leibniz, and much later, information-processing models in cognitive psychology, Noam Chomsky's theory of generative grammar, and programs for language and reasoning in artificial intelligence.

But there is an alternative:

"There appear to be only three principles of connection among ideas, namely, *resemblance*, *contiguity* in time or place, and *cause or effect*. Experience teaches us that a number of uniform effects result from certain objects. When a new object, endowed with similar sensible qualities, is produced, we expect similar powers and forces and look for a like effect. From a body

of like color and consistence with bread we expect like nourishment and support."



In this passage from his 1748 *Enquiry Concerning Human Understanding*, David Hume summarizes the theory of *associationism*. The mind connects things that are experienced together or that look alike, and generalizes to new objects according to their resemblance to known ones. Replace Hume's "ideas" or "sensible qualities" with "stimuli" and "responses," and you get the behaviorism of Ivan Pavlov, John Watson, and B. F. Skinner. Replace the ideas with "neurons" and the associations with "connections," and you get the neural network models of D. O. Hebb and the school of cognitive science called connectionism.

The theories would not have survived for centuries if they did not account for important phenomena. Associationism captures the tendency of animals to pick up statistical patterns among events and generalize them to similar events. Examples range from the gradient of bar-pressing rates in rats when the surrounding stimuli vary from training conditions to the widely reported demonstration in these pages last year that eight-month-old infants pick up the probabilities of transition between syllables in streams of artificial speech (3).

Moreover, it's easy to see how the laws of association might be implemented in neural hardware. If, as many neuroscientists believe, neurons that fire together wire together, we have an implementation of Hume's principle of contiguity in time. If neurons represent simple properties, and sets of active neurons represent concepts, then concepts that are similar will literally overlap in neural real estate, and anything associated with one concept will automatically be

associated with similar concepts. The connectionists Geoffrey Hinton, David Rumelhart, and James McClelland, echoing Hume's remark about resemblance, wrote, "If ... you learn that chimpanzees like onions you will probably raise your estimate of the probability that gorillas like onions. In a network that uses distributed representations, this kind of generalization is automatic" (4).

The theory of symbol processing seems better suited to explaining the brain's ability to handle complex ideas and the aspects of language that communicate them. People are not slaves to similarity. We can be told that a whale is not a fish and that Tina Turner is a grandmother, overriding our statistical experience of what fish and

grandmothers tend to look like. This suggests an ability to summarize an entire category by a mental variable or symbol, whose meaning comes from the rules it enters into: "a mammal is an animal that suckles," "a grandmother is the mother of a parent." These rules support generalizations that work more like deductions than similarity gradients. For example, we can infer that whales have livers or that Ms. Turner has had at least one baby (5).

Language is the quintessential symbol-manipulating system. When we learn that the grammatical object comes after the verb from simple sentences like "Tex hugged the dog," we can generalize that regularity to grammatical objects that are very different in sound ("I like Joe Bfbsplk"), in meaning ("Kant defined the categorical imperative"), or in length ("Sheila met a tall blonde man with one brown shoe"). The abstractness and open-ended expressive power of human language comes from a system of recursive rules manipulating variables like "noun phrase" and "object" (6).

Although many cognitive scientists believe that the human mind is a hybrid system that uses both associations and rules (5), others want to retain associative networks as the fundamental stuff of cognition (4). They suggest that humans are not naturally good at the kind of reasoning subserved by rules. Rule use emerges late in life as a result of formal schooling and socially articulated rules, or as a result of extensive training that makes an associative network approximate rule-like behavior.

Marcus *et al.* (1) have now shown that infants as young as seven months can abstract simple rules from language-like sounds, suggesting that rule formation is not a late add-on but there from the start.

Children of that age are just beginning to segment words from ambient speech, although they are several months away from understanding or producing them (6). Marcus *et al.* used a common method in the study of infant cognition: present a stimulus repeatedly until the infants are bored, then present them either with stimuli of the same kind or of a different kind. "Same kind" and "different kind" are in the mind of the beholder, so if infants attend longer to the different kind, they must be telling them apart.

In these experiments, infants were habituated with "sentences" that follow one sequence, such as "ga ti ga" and "li na li" (an ABA pattern), and then were presented with sentences that contained different words and either the same sequence, such as "wo fe wo" (ABA), or a different sequence, such as "wo fe fe" (ABB). The babies listened longer to the "different" sequence, showing that they must have discriminated ABA from ABB; everything else about the test sentences, such as the actual syllables and their transition probabilities, was the same. Various controls ensured that the children did not simply like the sound of some sequences more than others, or memorize smaller chunks like BA. Marcus has also demonstrated that a kind of associative network frequently touted as a ruleless model of language learning, J. Elman's Simple Recurrent Network, does not discriminate the patterns in the way these infants do.

Marcus *et al.* (1) are careful not to claim that infants lack an ability to form associations, that rule learning is uniquely human, or that the rule-learning mechanism at work in this experiment is the same one that babies use to acquire language later. But their demonstration suggests that the ability to recognize abstract patterns of stimuli that cut across their sensory content is a basic ability of the human mind. How it is carried out in the brain is still largely a mystery. Research in the neurobiology of learning and in neural network modeling (perhaps searching where the light is best) has tended to focus on simple associative learning mechanisms whose functions would have been recognizable to associationist philosophers writing centuries ago. Marcus *et al.*'s experiment is a reminder that humans also think in abstractions, rules, and variables, and is a challenge to figure out how we do so.

References and Notes

1. G. F. Marcus, S. Vijayan, S. Bandi Rao, P. M. Vishton, *Science* **283**, 77 (1999).
2. T. Hobbes, *Leviathan* (1651; reprint, Oxford Univ. Press, New York, 1957).
3. J. Saffran, R. Aslin, E. Newport, *Science* **274**, 1926 (1996).
4. G. E. Hinton, J. L. McClelland, D. E. Rumelhart, in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. (MIT Press, Cambridge, MA, 1986).
5. S. L. Armstrong *et al.*, *Cognition* **13**, 263 (1983); S. Pinker, *How the Mind Works* (Norton, New York, 1997); G. F. Marcus, *The Algebraic Mind* (MIT Press, Cambridge, MA, 1999).
6. S. Pinker, *The Language Instinct* (HarperCollins, New York, 1994).
7. Grant support received from NIH HD 18381.

The author is in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: steve@psyche.mit.edu