Modeling infant learning via symbolic structural alignment

Sven E. Kuehne (skuehne@ils.nwu.edu)
Department of Computer Science, Northwestern University
1890 Maple Avenue, Evanston, IL 60201 USA

Dedre Gentner (gentner@nwu.edu)
Department of Psychology, Northwestern University
2029 Sheridan Rd., Evanston, IL 60201 USA

Kenneth D. Forbus (forbus@ils.nwu.edu)
Department of Computer Science, Northwestern University
1890 Maple Avenue, Evanston, IL 60201 USA

Abstract

Understanding the mechanisms of learning is one of the central questions of Cognitive Science. Recently Marcus et al. showed that seven-month-old infants can learn to recognize regularities in simple language-like stimuli. Marcus proposed that these results could not be modeled via existing connectionist systems, and that such learning requires infants to be constructing rules containing algebraic variables. This paper proposes a third possibility: that such learning can be explained via structural alignment processes operating over structured representations. We demonstrate the plausibility of this approach by describing a simulation, built out of previously tested models of symbolic similarity processing, that models the Marcus data. Unlike existing connectionist simulations, our model learns within the span of stimuli presented to the infants and does not require supervision. It can handle input with and without noise. Contrary to Marcus' proposal, our model does not require the introduction of variables. It incrementally abstracts structural regularities, which do not need to be fully abstract rules for the phenomenon to appear. Our model also proposes a processing explanation for why infants attend longer to the novel stimuli. We describe our model and the simulation results and discuss the role of structural alignment in the development of abstract patterns and rules.

Introduction

Understanding the mechanisms of learning is one of the central questions of cognitive science. Recent studies (Gomez & Gerken, 1999; Marcus, Vijayan, Rao & Vishton, 1999) have shown that showed that infants as young as seven months can process simple language-like stimuli and build generalizations sufficient to distinguish familiar from unfamiliar patterns in novel test stimuli. In Marcus et al's study, the stimuli were simple 'sentences,' each consisting of three nonsense consonant-vowel 'words' (e.g., 'ba', 'go', 'ka'). All habituation stimuli had a shared grammar, either ABA or ABB. In ABA-type stimuli the first and the third word are the same: e.g, 'pa-ti-pa.' In ABB-type stimuli the second and the third word are identical: e.g., 'le-di-di'. The infants were habituated on 16 such sentences, with three repetitions for each sentence. The infants were then tested on a different set of sentences that consisted of entirely new words. Half of the test stimuli followed the same grammar as in the habituation phase; the other half followed the non-trained grammar. Marcus et al. found that the infants dishabituated significantly more often to sentences in the non-trained pattern than to sentences in the trained pattern.

Based on these findings Marcus et al. proposed that infants had learned abstract algebraic rules. They noted that these results cannot be accounted for solely by statistical mechanisms that track transitional probabilities. They further argue that their results challenge connectionist models of human learning that use similar information, on two grounds: (1) the infants learn in many fewer trials than are typically needed by connectionist learning systems; (2) more importantly, the infants learn without feedback. In particular, Marcus et al. demonstrated that a simple recurrent network with the same input stimuli could not model this learning task.

In response, several connectionist models have attempted to simulate these findings. Unfortunately, all of them to date include extra assumptions that make them a relatively poor fit for the Marcus et al experiment. For example, Elman (1999; Seidenberg & Elman, 1999) use massive pre-training (50,000 trials) to teach the network the individual stimuli. More importantly, they turn the infants' unsupervised learning task into a supervised learning task by providing the network with external training signals. Other models tailored to capture the data of the study seem unlikely to be applicable to other similar cognitive tasks (Altmann & Dienes, 1999). Using a localist temporal binding scheme, Shastri and Chang (1999) model the infant results without pretraining and without supervision, but still require an order of magnitude more exposure to the stimuli than the infants received.

We propose a third alternative. There is evidence that structural alignment processes operating over symbolic structured representations participate in a number of cognitive processes, including analogy and similarity (Gentner, 1983), categorization (Markman & Gentner, 1993), detection of symmetry and regularity (Ferguson, 1994), and learn-

ing and transfer (Gentner & Medina, 1998). Although these representations and processes are symbolic, they do not need to be rule-like, nor need they involve variables. Instead, we view the notion of correspondence in structural alignment as an interesting cognitive precursor to the notion of variable binding[1]. Correspondences between structured representations can support the projection of inferences, as the analogy literature shows, and therefore a symbolic system can draw inferences about novel situations even without having constructed rules. Moreover, as discussed below, comparison can be used to construct conservative generalizations. Across a series of items with common structure such a process of progressive abstraction can eventually lead to abstract rule-like knowledge. The attainment of rules, in those cases where it occurs, is the result of a gradual process. As we will show, symbolic descriptions can be used with structural alignment to model learning that is initially conservative, but which occurs fast enough to be psychologically realistic.

We first describe our simulation model of the Marcus et al task, which uses a simple combination of preexisting simulation modules, i.e., SME, MAGI, and SEQL. All of these modules have been independently tested against psychological data and independently motivated in prior modeling work. With the exception of domain-specific encoding procedures, no new processing components were created for this task. We then describe the results of our simulation of the Marcus et al data, showing that our simulation can learn the concepts within the number of trials that the infants had, without supervision and without pre-learning. We also show that the simulation can exhibit the same results with noisy input data. Finally, we discuss some of the implications of the symbolic similarity approach for models of cognitive processing.

Modeling infant learning via structural alignment

A psychological model of the infants' learning must include the kind of input, the way the infants are assumed to encode the individual sentences, and the processes by which they generalize across the sentences. The architecture of our simulation is shown in Figure 1. We first describe our assumptions concerning the infants' processing capacities. Then we describe each component in turn.

Processing Assumptions: We assume that infants can represent the temporal order within the sentences (Saffran, Aslin & Newport, 1996). We further assume that the infants notice and encode identities within the sentences: for example, the fact that the last two elements match in an ABB sentence. This assumption is consistent with evidence that human infants, as well as with studies of nonhuman primates (Oden et al, in press), can detect identities. We also assume that infants can detect similarities between sequentially presented stimuli, consistent with studies of infant habituation, which demonstrate that infants respond to sequential sameness (e.g., Baillargeon, 1994).
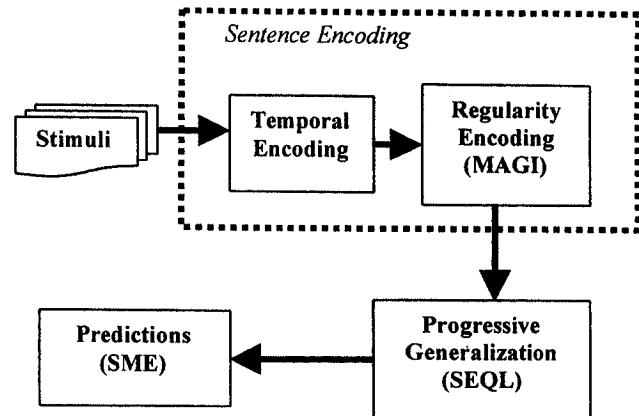
---

[1] That structure-mapping algorithm neither subsumes, nor is subsumed by, traditional pattern matching such as unification is shown in Falkenhainer, Forbus, & Gentner (1988).

Figure 1: Simulation Architecture

Input stimuli: To make our simulation comparable with others, we use a representation similar to that of Elman (1999), namely, Plunkett & Marchman's (1993) distinctive feature notation. Each word has twelve phonetic features, which can be either present or absent. The presence or absence of each feature for each word is encoded by symbolic assertions. If feature n is present for word w, the assertion $(Rn\ w)$ is included in the stimulus, and if absent, the assertion $(Sn\ w)$ is included. Thus the acoustic features of each word are encoded as twelve attribute statements.

We modeled the Marcus et al experiment both without noise (Experiment 1) and with noise (Experiment 2). Marcus et al. used a speech synthesizer to control the pronunciation of the stimuli, but while this reduces variability, it cannot eliminate the possibility that the infant might encode something incorrectly.

Temporal encoding: We assume that the infant encodes the temporal sequence of the words in a sentence in two ways. First, each incoming word has an attribute associated with it, corresponding to the order in which it appears (i.e., FIRST, SECOND, or THIRD). We further assume that the infant encodes temporal relationships between the words in a sentence:; to code this, an AFTER relation is added between pairs of words in the same sentence indicating their relative temporal ordering. The particular labels used in this encoding step are irrelevant – there are no rules in the system that operate on these specific predicates – the point is simply that infants are encoding the temporal order of words within sentences.

Regularity Encoding: We assume that the infants notice and encode identities within the sentences: for example, the fact that the last two elements match in an ABB sentence. Thus the simulation must incorporate a process that detects when words are the same. We use the MAGI model of symmetry and regularity detection (Ferguson, 1994) to automatically compute these relationships. MAGI treats symmetry as a kind of self-similarity, using a modified version of structure-mapping's constraints to guide the self-alignment process. MAGI has been successfully used with inputs ranging from stories to mathematical equations to visual stimuli,

and it has done well at modeling certain aspects of visual symmetry, including making new predictions (Ferguson *et al* 1996). Here MAGI is used on the collection of words in a sentence. For any pair of words *w1* and *w2* that MAGI finds sufficiently similar, this module asserts (SIM *w1 w2*), and a DIFF statement for every other pair of words in the sentence. (If MAGI does not find any pairs similar, DIFF statements are asserted for every pair of words.) This module also asserts (GROUP *w1 w2*) for pairs of similar words, to mark that they form a substructure in the stimulus, and adds DIFF statements between groups and words not in the group. This use of MAGI is an example of what Ferguson (1994, in preparation) calls *analogical encoding*.

SEQL

Once each sentence is encoded, we assume infants can detect the similarities between sequential pairs of sentences. The detection of structurally parallel patterns across a sequence of examples is modeled by SEQL (Skorstad, Gentner & Medin, 1988; Kuehne, Forbus, Gentner & Quinn, 2000), a model of the process of category learning from examples. SEQL constructs category descriptions via incremental abstraction. That is, the representation of a category is a structured description that has been generated by successive comparison with incoming exemplars. If the new exemplar and the category are sufficiently similar, the category description is modified to be their intersection -- i.e., the commonalities computed via structural alignment by a generalization algorithm. If the new exemplar is not sufficiently similar, it is stored separately and may later be used as the seed of a new category.

The structural alignment process is implemented via SME, (Falkenhainer *et al* 1988; Forbus *et al* 1994) a cognitive simulation of analogical matching. Here the base description is a category description, and the target description is the new exemplar. The structural alignments that SME computes are used in three ways by SEQL. First, the numerical structural evaluation score it computes[2] is used as a similarity metric, a numerical measure for deciding whether or not two descriptions are sufficiently similar. Second, the candidate inferences it computes serve as a model for category-based induction (c.f. Blok & Gentner, 2000; Forbus, Gentner, Everett, & Wu, 1997). Third, the correspondences in the best mapping SME produces serves as the basis for SEQL's generalization algorithm.

SEQL maintains a set of generalizations and a set of singular exemplars. When a new exemplar comes in, it is compared against existing generalizations to see if it can be assimilated into one of them. Otherwise, it is compared with the stored exemplars to see if a new generalization can be formed. If it is insufficiently similar to both the generalizations and the stored exemplars, it is stored as an exemplar itself.

SEQL begins with no generalizations; it simply stores its first exemplar. If the next exemplar is sufficiently close to the first, their overlap is stored as the first generalization. A

generalization consists of the overlap between the two input descriptions: that is, the shared structure found by alignment. Thus generalizations are structured descriptions of the same type as the input descriptions, although containing fewer specific features. If a new exemplar is sufficiently similar to a generalization (as determined comparing the structural evaluation score to a set threshold), then (a) the generalization is updated by retaining only the overlapping description that forms the alignment between the generalization and the exemplar; and (b) candidate inferences are projected from the generalization to the exemplar. Non-overlapping aspects of a description (e.g., phonetic features or relations that aren't shared) are thus "worn away" with each new assimilated description. (The threshold that determines when descriptions are sufficiently similar to be assimilated helps prevent descriptions from diminishing into vacuity.)

Returning now to the infant studies, we assume that babies are carrying out an ongoing process of comparing and aligning the incoming exemplars with an evolving generalization. We further assume that the relational candidate inferences from the general pattern to a new exemplar represent expectations on part of the infant.[3] When these expectations are violated by an incoming stimulus that does not fit the generalized pattern (e.g., an ABB test sentence after the ABA generalization has been formed), we assume the infant requires extra time to process the inconsistent stimulus.

Simulation Experiments

In both experiments, we followed the procedure of Marcus et al. Each stimulus was a simple three-word sentence, encoded as described earlier. There were two sets of training stimuli, one following the ABA pattern and one following the ABB pattern. The training stimuli were (ABA) de-di-de, de-je-de, de-li-de, de-we-de, ji-di-ji, ji-je-ji, ji-li-ji, ji-we-ji, le-di-le, le-je-le, le-li-le, le-we-le, wi-di-wi, wi-je-wi, wi-li-wi, wi-we-wi and (ABB) de-di-di, de-je-je, de-li-li, de-we-we, ji-di-di, ji-je-je, ji-li-li, ji-we-we, le-di-di, le-je-je, le-li-li, le-we-we, wi-di-di, wi-je-je, wi-li-li, wi-we-we. The test stimuli in both experiments were four descriptions representing two novel ABA-type (ba-po-ba, ko-ga-ko) and two novel ABB-type sentences (ba-po-po, ko-ga-ga). The threshold value for SEQL was set to 0.85 in both experiments.

Experiment 1

This experiment is most comparable to previous simulation models of the phenomena, in that we assume noise-free encoding of the stimuli. A simulation run consists of exposing SEQL to all of the stimuli from a particular training set (either ABA or ABB) once and then seeing the response given the four test sentences. To avoid possible biasing due to sequence effects (See Kuehne et al., 2000), 20 simulation runs were made for each training set using different random

---

[2] Although SME can compute multiple mappings, we use the structural evaluation score of the best mapping, normalized by the size of the base description.

[3] SME can also produce attribute-level candidate inferences, and does so on these stimuli. We assume that, since these inferences concern directly perceivable features, testing them takes very little time.

orders. Identical match score and relational candidate inferences were produced for all sequences with a given stimulus set. In each case, SEQL produced a single generalization during the learning phase. For the test phase we used encodings of the corresponding stimuli used with infants, as noted above. Tables 1a and 1b show the results of this series for two generalizations paired against the four test sentences.

Table 1a: ABA training stimuli

Test Stimulus	Match Score	Candidate Inferences
Ba-po-ba		None
Ko-ga-ko	.	None
Ba-po-po	0.486	(DIFF po1 ba1) (DIFF po1 po2) (SIM ba1 po2)
Ko-ga-ga	0.455	(DIFF ga1 ko1) (DIFF ga1 ga2) (SIM ko1 ga1)

Table 1b: ABB training stimuli

Test Stimulus	Match Score	Candidate Inferences
Ba-po-ba	0.328	(SIM po1 ba2) (DIFF ba1 (GROUP po1 ba2))
Ko-ga-ko	0.350	(SIM ga1 ko2) (DIFF ko1 (GROUP ga1 ko2))
Ba-po-po	.	None
Ko-ga-ga	.	None

The in-grammar (bold) and out-of-grammar (plain text) matches show clear differences in their match scores. In-grammar matches are above 0.64 and do not generate relational candidate inferences. Out-of-grammar matches have match scores below 0.5, and lead to relational candidate inferences. Thus out-of-grammar test sentences lead to longer looking behavior, as predicted.

Experiment 2

As noted earlier, we believe that noise-free stimulus encodings are unrealistic. Consequently, we used the same procedure as Experiment 1, but this time introducing noise into the representations for the training and test stimuli. For each sentence, one of the words was randomly picked, and one of its attributes (also chosen at random) was dropped or flipped, with the rest of its description being unchanged. Such changes can be significant: for example, flipping a single phonetic feature turns the word 'de' into the word 'di'. Again, 20 simulation runs were made for each training set using different random orders. Naturally the match scores and, to a lesser degree, the generated candidate inferences, did vary across the individual runs. Tables 2a and 2b show the results. The scores were averaged over all 20 runs.

Although the noise affected the details of the computations, the overall pattern of results remains the same. The in-grammar (bold) match scores are far higher than the out-of-grammar (plain text) scores; and the out-of-grammar

stimuli produce relational candidate inferences while the in-grammar stimuli do not.

Table 2a: ABA training stimuli

Test Stimulus	Average Match Score	Candidate Inferences Min, Average, Max
ba-po-ba	.	0, 0, 0
ko-ga-ko	0. 2	0, 0, 0
ba-po-po	0.435	2, 2.45, 3
ko-ga-ga	0.395	2, 2.55, 3

Table 2b: ABB training stimuli

Test Stimulus	Match Score	Candidate Inferences Min, Average, Max
ba-po-ba	0.339	2, 2, 2
ko-ga-ko	0.352	2, 2.05, 3
ba-po-po	0. 0	0, 0, 0
ko-ga-ga	0.	0, 0, 0

Comparison with other models

The results of Marcus et al. (1999) have sparked an active debate focused on two issues: (1) Can current connectionist models (e.g., simple recurrent networks) model these results? (2) Do infants generate abstract rules that include variables?

Regarding the adequacy of simple recurrent networks, Marcus et al. state "Such networks can simulate knowledge of grammatical rules only by being consequently 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 the training." Elman's (1999) response describes his use of a simple recurrent network to model this task. Elman's model requires tens of thousands of training trials on the individual syllables, and treats the problem as a supervised learning task, unlike the task facing the infants. By contrast, our simulation handles the learning task unsupervised, and produces human-like results with only exposure to stimuli equivalent to that given to the infants. Moreover, our model also continues to work with noisy data, something not true of any other published model of this phenomenon that we know of.

The learning in our model is due to the "wearing away" of non-identical phonetic attributes through subsequent comparisons. Although SEQL's learning proceeds faster than connectionist models, it is still slower than systems that generate abstractions immediately (e.g., explanation-based learning (DeJong & Mooney, 1986)). In SEQL's progressive alignment algorithm, the entities in the generalizations lose their concrete attributes across multiple comparisons, leaving the relational pattern of each grammar as the dominant force in the generalization only after a reasonable num-

ber of varied examples are seen.[4] There is considerable evidence for this kind of conservative learning (Forbus & Gentner, 1986; Medin & Ross, 1989).

Turning to the second issue, whether infants have variables and generate abstract rules, Marcus et al (1999) claims "[I]nfants extract abstract algebra-like rules that represents 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.'" But our simulation does not introduce variables, in the sense commonly used in mathematics or logic. The generalizations constructed by SEQL do indeed include relational patterns that survive repeated comparisons because they are shared across the ingrammar exemplars. Furthermore, the entities (words) in the generalizations have many fewer features than the original words, as a result of the wearing away of features in successive comparisons. One could consider these patterns as a form of psychological rule, as proposed by Gentner and Medina (1998), with the proviso that the elements in the rule are not fully abstract variables, although they might asymptotically approach pure variables.

Discussion

This paper proposes a third kind of explanation for the infant learning phenomena of Marcus et al (1999): incremental abstraction of symbolic descriptions via structural alignment. We believe our explanation is currently the best one for three reasons. First, it models the infant data with fewer extra concessions than previously published models (i.e., no pre-training, no supervision, and noisy data). Second, the processes we postulate are cognitively general; they apply to a large set of phenomena. Third, the abstraction processes we propose are consistent with research demonstrating that human learning is initially conservative (Brooks, 1987; Forbus & Gentner, 1986; Medin & Ross, 1989). Interestingly, there is ongoing research in developing symbolic connectionist models consistent with these processes (e.g., Holyoak & Hummel, 1997).

Many issues remain to be explored. For example, although our system does not introduce variables in its generalization process, there is a sense in which the entities in the generalization are on their way to becoming variables. Gentner and Medina (1998) have proposed that the process of progressive alignment can lead to rules. They further suggested that the application of rules to instances can be accomplished using the same general processes of structural alignment and projection that are used in analogy. The difference is that the base domain is an abstraction, the entities are 'dummies' with no features to either help or impede the match with the specific entities in the exemplar. Another issue concerns the incorporation of statistical notions in SEQL. Although SEQL is to a certain degree noise-resistant,

we suspect that to model large-scale learning, it will need to keep track of more statistical information than it does currently, so that properties wear away more slowly.

We note that it is common to conflate symbolic processing with rule-based behavior, and parallel processing with connectionist models. The model described here is symbolic, but it need not involve variables or rules. Further, it involves extensive parallel processing (most of SME and MAGI's computations are parallel). Given the complexity of the phenomena, such conflations seem unwise.

The debates stirred by the Marcus et al. results bear on a critical issue in human learning and development: namely, what knowledge or mechanisms must be assumed to account for the rapid and powerful achievements demonstrated by infants in both cognition and language. Our results suggest that the general learning mechanism of structure-mapping theory may go a long way in accounting for these accomplishments.

References

Altmann, G.T.M. and Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks, *Science 284*, 875.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science, 3*(5), 133-140.

Blok, S. V., & Gentner, D. (2000). Reasoning from shared structure. *Proceedings of the 22nd Meeting of the Cognitive Science Society.*

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: The ecological and intellectual factors in categorization* (pp. 141-174). Cambridge: Cambridge University Press.

Christiansen, M.H. and Curtin, S.L. (1999). The power of statistical learning: No need for algebraic rules, in Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society, Erlbaum, Mahway, NJ.

Christiansen, M.H. and Curtin, S.L. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends in Cognitive Science 3*, 289-290

DeJong, G.F. and Mooney, R.J. (1986). Explanation-based learning: An alternative view. *Machine Learning 1*(2), pp. 145-176

Elman, J. (1999). Generalization, rules, and neural networks: A simulation of Marcus et. al, (1999). Ms., University of California, San Diego.

Falkenhainer, B., Forbus, K., and Gentner, D. (1986). The Structure-Mapping Engine. In: *Proceedings of AAAI 86*, Philadelphia, PA, August.

---

[4] SEQL learns with only one exposure to the 16 learning sentences, whereas Marcus's infants received three exposures for each sentence. It is possible that the infants would have learned with only one pass; however it is also possible that the infants were less consistent in detecting the similarities than our simulation with its current parameters.

Falkenhainer, B., Forbus, K.D. and Gentner, D. (1989). The Structure Mapping Engine: an algorithm and examples. *Artificial Intelligence, 41*: 1-63

Ferguson, R.W. (1994). MAGI: A model of analogical encoding using symmetry and regularity. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Ferguson, R.W., Aminoff, A. and Gentner, D. (1996). Modeling qualitative differences in symmetry judgments. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Forbus, K. D., & Gentner, D. (1986). Learning physical domains: Toward a theoretical framework. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 311-348). Los Altos, CA: Kaufmann.

Forbus, K. D., Ferguson, R. W., and Gentner, D. (1994). Incremental Structure-mapping. In: *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science, 7*: 155-170.

Gentner, D. and Markman, A.B. (1997). Structure-mapping in analogy and similarity. *American Psychologist, 52,* 45-56.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263-297.

Goldstone, R.L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition 52*(2), 125-157.

Goldstone, R.L., Medin, D.L., and Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgements. *Cognitive Psychology, 23,* 22-264.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition 70,*109-135.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*(3), 427-466.

Kuehne, S.E., Forbus, K.D., Gentner, D. and Quinn, B. (2000). SEQL- Category learning as incremental abstraction using structure mapping, *Proceedings of the Twenty-second meeting of the Cognitive Science Society.*

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

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

Marcus, G.F. (1999). Simple recurrent networks and rule-learning: http://psych.nyu.edu/~gary/science/es.html.

Markman, A.B. and Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.

McClelland, J.L. and Plaut, D.C. (1999). Does generalization in infant learning implicate abstract algebraic rules?, *Trends in Cognitive Science 3,* 166-168

Medin, D.L., Goldstone, R., and Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254-278.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189-223). Hillsdale, NJ: Erlbaum.

Oden, D. L.. Thompson, R. K. R., and Premack, D. (in press). Can an ape reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee (Pan troglodytes). In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspecitives from cognitive science.* Cambridge, MA: MIT.

Plunkett, K. and Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21-69.

Saffran, J., Aslin, R. and Newport, E. (1996). Statistical learning by 8-month-old infants, *Science, 274,* 1926-1928

Seidenberg, M.S. and Elman, J. (1999), Do infants learn grammar with algebra or statistics?, Letter, *Science 284,* 434-436

Seidenberg, M.S. and Elman, J. (1999). Networks are not hidden rules, *Trends in Cognitive Science 3,* 288-289

Skorstad, J., Gentner, D. and Medin, D. (1988). Abstraction processes during concept learning: a structural view. In: *Proceedings of the Tenth Annual Conference of the Cognitive Science Society.* Montreal: Lawrence Erlbaum Associates.