

# The Role of “Explaining Away” in Human Abstract Rule Induction

**Colin Reimer Dawson, M.A. (CDawson@email.arizona.edu)**

Department of Psychology, 1503 E. University Blvd.  
Tucson, AZ 85721 USA

**LouAnn Gerken, Ph.D. (Gerken@email.arizona.edu)**

Department of Psychology, 1503 E. University Blvd.  
Tucson, AZ 85721 USA

## Abstract

Of great interest to cognitive science is how human learning is constrained to avoid spurious generalizations. While many constraints must be relatively experience-independent, past experience provides a rich source of guidance for subsequent learning. If a learner discovers some structure in part of the environment, this can inform her future hypotheses about that domain. If a general structure parsimoniously accounts for particular sub-patterns, a rational learner should not stipulate separate explanations for each detail without additional evidence, as the general structure has “explained away” the original evidence. In a grammar-learning experiment using tone sequences, manipulating learners’ prior exposure to a tone environment affects their sensitivity to the grammar-defining feature, in this case consecutive repeated tones. Grammar-learning performance is worse if context melodies are “smooth”, that is, if small intervals occur more often than large ones, as this smoothness is a general property that accounts for a high rate of repetition.

**Keywords:** statistical learning; artificial grammar; Bayesian inference; language acquisition; music cognition

## Introduction

In traditional theories of learning, the relationship between knowledge<sup>1</sup> and learning is fairly static. Some initial knowledge is provided by experience-invariant biology to constrain learning. Within these *a priori* constraints, learning builds one’s body of knowledge. An area that has been explored relatively little is the dynamic interplay between learning and knowledge: namely, (how) can the results of learning actually change how subsequent learning proceeds? If this feedback loop is ignored, observed constraints on learning may be incorrectly attributed to initial biologically provided knowledge instead of to learning that has already taken place.

Untangling the relative contributions of experience-independent biology and prior learning has been particularly important in the study of infant cognition, not least infant language acquisition. If an adult can learn one pattern and not another in the absence of *a priori* differences in difficulty, there are often ready explanations in terms of her years of experience in the world. In contrast, if a young infant exhibits the same discrepancy, it is tempting to

attribute it to biology. This conclusion would be premature without further examination, however.

Indeed, previous research suggests that infants reorganize their domain knowledge within the first year of life, and even within the laboratory. In language, infants reorganize their phonetic categories (Werker & Tees, 1984; Bosch and Sebastian-Galles, 2003; Maye, Werker, & Gerken, 2002) and even exhibit shifts in what features a stress rule can reference (Gerken and Bollt, 2008). In music, attention shifts from absolute pitches to relative intervals (Saffran and Griepentrog, 2001; Saffran, 2003), and infants’ tonal and rhythmic categories change as a function of cultural context (Hannon and Trehab, 2005; Lynch and Eilers, 1992).

Marcus, et al. (1999) and Marcus, Fernandes and Johnson (2007) found that 7-month-old infants can learn an AAB or ABB pattern in three-element sequences, provided the elements are syllables. Infants at the same age failed at the same task when the elements were non-linguistic events such as musical tones or animal noises. It was suggested that the child’s initial endowment may tell her that speech can be structured by abstract, relational properties, but other auditory stimuli cannot.

While this is possible, subsequent research has revealed AAB-style learning in infants with other stimuli, such as pictures of dogs (Saffran, et al., 2007), and simple shapes (Johnson, et al., in press). Murphy, Mondragon and Murphy (2008) found that even rats can learn such generalizations from both speech and tones. These results cast doubt on the notion that language is privileged for abstract pattern-learning.

## “Explaining Away” Details With Generalities

Dawson and Gerken (2009) found that while 7-month-olds fail at learning AAB and ABA patterns with tones, 4-month-olds succeed given the same input. They suggested that 7-month-olds’ failure may be due to their having learned certain general properties about music. In particular, if they have learned that (a) melodies tend to move in small intervals from pitch to pitch, and (b) individual melodies tend to use only a restricted set of pitches (Temperley, 2008), the presence of a large number of repetitions would become much less surprising, and hence less informative about the abstract structure in the AAB-style task. This change in informativeness is an example of a phenomenon known as “explaining away”, central to several cognitive models in a variety of areas

<sup>1</sup> Here, “knowledge” is meant in a broad sense – roughly, “information about the environment”. This can be anything that affects behavior, or, critically, the interpretation of experiences.

including visual inference (Kersten, Mamassian and Yuille, 2004), linguistic processing (Ciaramita and Johnson. 2000), and infant causal reasoning (Xu and Garcia, 2008; Gergely and Csibra, 2003).

The basic idea is as follows. When an observed pattern could arise from multiple hidden causes, the causes “compete” with each other over the evidence contained in the data, even when the underlying hypotheses do not conflict with each other *a priori*. For example, suppose during a card game you peek at the dealer’s hand, and you notice that on one hand, she has three aces, and on the next she has the nine through king of hearts. If you assume the game is poker, this unusually lucky sequence might raise suspicion that the dealer has stacked the deck to give herself a favorable hand. However, if you later learn that the players are engaged in a friendly pinochle match, in which only the cards nine through Ace are used, the dealer’s hands are less surprising given a fair deal. Although the dealer may still be stacking the deck, the evidence for this hypothesis must be discounted, or “explained away”.

In a musical context, repetition is an ambiguous event. On the one hand, it constitutes a “sameness” relation between two tones. At the same time, it is also an interval of magnitude zero between successive pitches. If one assumes that melodies are random, and that any tone is equally likely at any point (i.e., the tone distribution is uniform), hearing every melody begin with two repeated notes would be quite surprising, and evidence for a “sameness” interpretation would be strong. If, however, one knows that tones nearby in time also tend to be nearby in pitch (i.e., melodies are usually “smooth”), repetition becomes a more common event (*qua* interval of distance zero), and it should take more evidence to conclude that repetition is special. Similarly, as the set of tones shrinks, the probability of chance repetitions increases (as with the three aces in the Pinochle hand), and the evidentiary bar for learning a repetition grammar should be raised.

The present experiment provides a test of the first of these two predictions with human adults. Participants are first placed in one of three melodic environments: one where every tone is equally likely at any point (the Uniform condition); one in which small intervals are more common than large intervals (the Smooth condition); and one in which repetition alone is more frequent than other intervals (the Repetition condition). Following this exposure, participants are given a grammar-induction task where the “grammatical” melodies have either an AABCD or DCBAA structure. If learners model the interval distribution in the larger environment, the Smooth context should lead them to represent repetition as the result of a general constraint on melodies, and not as a specific grammatical feature. Hence, learners should exhibit decreased sensitivity to positional repetition, as well as decreased grammar-learning performance.

In contrast, in the Repetition environment, the only way to explain the high rate of repeated tones is to represent it explicitly. This unexplained repetition may even *increase*

learners’ attention to that feature, improving their performance relative to the Uniform group.

## Methods

### Participants

One hundred and twenty University of Arizona undergraduates participated in the study for course credit. An additional eighteen participated but were excluded from analysis due to their failure to score above chance on a melodic-discrimination screening task.

### Materials and Procedures

The experiment consists of a “context” phase and a grammar-learning phase. The latter contains four blocks, each with a training component and a test component. All “sentences” consist of five tones generated using the FM Synthesizer in the MIDI Toolbox for MATLAB (Eerola & Toivainen, 2004), which produces a horn-like sound. The first four notes are 250 msec each, with 50 msec gaps after each one. The last note is 500 msec. In music terms, the melodies contain four eighth notes followed by a quarter note, played at 200 beats per minute.

**Procedures: Context Phase** The context phase consists of two blocks of 100 sentences, in random order. Ten are “probe” sentences, after which either the same sentence is repeated or one of the other ten probe sentences is played. On the probe trials, participants have 3 seconds to press the “1” or “0” key on the keyboard to register “same” or “different” sentence pairs. The absence of a response is coded as incorrect. Each block lasts about five minutes. Data from participants who did not perform above chance on this discrimination task (15 or more out of 20 correct) was discarded, as these participants presumably either could not distinguish differences among melodies, or were not attempting to succeed.

During context exposure, all participants see a group of eight cartoon “aliens” (Folstein, Van Petten and Rose, 2007). Half are “star-chested” and half are “brick-chested”.

**Materials: Context Phase** Participants are assigned to one of three context conditions: Uniform ( $n = 24$ ), Smooth ( $n = 48$ ) or Repetition ( $n = 48$ ). The Smooth and Repetition conditions are further divided into High Variance (HV) and Low Variance (LV) sub-conditions. In all cases, context melodies are drawn from a “vocabulary” of six tones: A3, A#3, C#4, E4, G4 and G#4 (MIDI values 57, 58, 61, 64, 67 and 68).

In the Uniform condition, each tone is equally likely and independent of the last. As such, the probability of a repetition at any given point is 1/6 (in the 200 generated melodies, the empirical rate was 18.1%). The resulting distribution of intervals is shown in Fig. 1a.

In the Smooth condition, melodies are generated as follows. The first tone is chosen from a uniform distribution over the six tones. For each subsequent tone, a

sample is generated from a normal distribution, truncated between 0.5 and 6.5. The mean of the distribution is an integer corresponding to the previous tone (the lowest tone is 1; the highest tone 6). The standard deviation is 2 in the HV condition and 1.2 in the LV condition. The sampled value is rounded to the nearest integer to generate the tone. The resulting distribution reflects the bias toward small intervals in typical folk music (Dawson, 2007). The rate of repetition across the 200 melodies is 39.3% of all intervals in the LV condition (Fig. 1b), and 26.3% in the HV condition (Fig. 1c).

The Repetition conditions control for the actual rate of repetition, while removing the overall “smoothness” constraint. Here, the HV and LV conditions (Fig. 1d-e) are matched to their Smooth counterparts for the number of repetitions, but unlike in the Smooth cases, the remaining notes are equiprobable. Here, the high rate of repetition cannot be explained by a general bias for small intervals; instead, a learner modeling the tone distribution must encode repetitions separately to achieve a good fit.

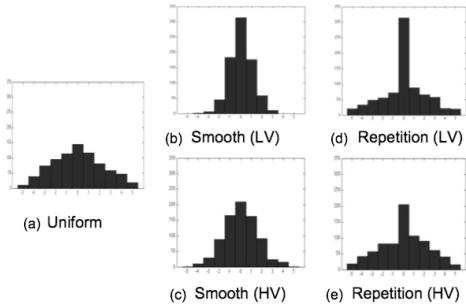


Figure 1(a-e): Interval Counts in the Context Phase<sup>2</sup>

**Procedures: Grammar-Learning Phase** After the context phase, participants move on to the grammar-learning phase. They are asked to detect “spies” attempting to infiltrate the “Qixian” colony, and are told that they can distinguish Qixians from spies by the grammaticality or ungrammaticality of their speech.

In each training block, participants hear thirty “grammatical” sentences in random order while an image of four star-chested aliens is displayed.

After each training block, participants hear twenty-four test sentences, half grammatical. After each sentence, participants make a continuous grammaticality judgment by clicking on a line (Fig. 2), where the left pole represents “definitely grammatical”, the right pole represents “definitely ungrammatical”, and every gradient response in between is possible. There is no time limit. The computer records a binary response, based on whether the participant

<sup>2</sup> Here, 0 is a repetition, +1 is a step to the next-highest note, etc. The distributions are combined across all pitches. The truncation of the pitch range results in more small intervals across all conditions. If the interval distributions were separated by preceding pitch, those for the Uniform and Repetition conditions would each be flat, except for the peak at 0 in the Repetition case.

clicks left or right of center, and a continuous “discrimination score” calculated by subtracting from 100 the percentage of the line lying between the response and the correct pole. Participants experience four training-test cycles on the same grammar.

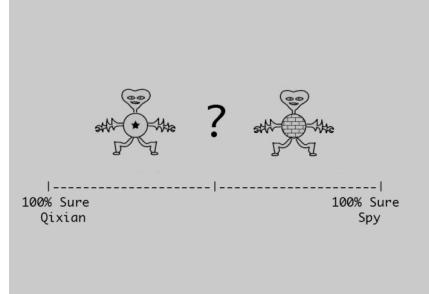


Figure 2: Test Prompt

**Materials: Grammar-Learning Phase** The “Qixian” and “spy” sentences are again five tones in length. Each participant is trained using one of two five-tone vocabularies. The first (V1) contains the tones A3, C4, D#4, F#4 and G4 (MIDI 57, 60, 63, 66 and 67); the second (V2) contains the tones A#3, B3, D4, F4 and G#4 (MIDI 58, 59, 62, 65 and 68). Each set shares two tones with the context vocabulary.

For half of participants, the “grammatical” sentences follow an AABCD pattern (with a repetition at the beginning and nowhere else), while the “ungrammatical” sentences have a DCBAA pattern. For the other half of participants, the labels are reversed.

Of the 120 sentences possible in each grammar, 60 are used as training items, and 24 as test items. The chosen items were balanced for pitch contour:  $\frac{1}{4}$  in each section had a rising segment followed by a falling segment (in addition to the repetition),  $\frac{1}{4}$  had the reverse;  $\frac{1}{4}$  had a rise-fall-rise pattern and  $\frac{1}{4}$  a fall-rise-fall pattern.

Thirty training sentences are used in the first two learning blocks; the other thirty in the last two blocks. On odd-numbered test blocks, participants are tested with items from the training vocabulary; on even blocks they hear items from the opposite vocabulary. Both vocabularies were used to test whether the context manipulation has an effect on the level of abstraction at which participants learn the grammar. The training vocabulary always comes first, as the vocabulary switch could provide a clue to the nature of the grammar (i.e., that it was vocabulary-independent), and if the new vocabulary came first, participants could not demonstrate mastery independent of this “hint”.

## Results

Of primary interest is whether prior exposure to the Smooth distribution will impair participants’ detection of the repetition pattern. If so, this will suggest that learners are establishing a higher baseline for repetition, which (partially) explains away the training pattern. The key comparison is between the Smooth and Repetition conditions, as these are matched for number of repetitions,

differing only in the presence or absence of a larger-scale regularity that accounts for that frequency.

A secondary question is whether the presence of an *inexplicably* high rate of repetitions will encourage learners to encode discrete “same” and “different” relations at the expense of the continuous relations among frequencies, thereby increasing the proportion of attention allocated to repetition and hence increasing performance in grammar-learning. If so, the Repetition group should outperform the Uniform group.

Pilot data revealed that many participants performed near ceiling at discriminating grammatical and ungrammatical sentences, while another large set performed at chance overall. For many of these, presumably only a fairly strong manipulation would observably shift performance. As such, the particular values of the scores received by these participants are mostly uninformative, and contribute noise that could obscure effects of the manipulations.

To address this issue, participants were separated into quartiles within each context condition based on their combined number of correct responses throughout the four test blocks, and two sets of analyses were conducted. The first used all of the data; the second discarded the highest- and lowest-performing quartiles in each condition, thereby greatly reducing the proportion of participants performing either at floor or ceiling. When “floor” is defined as producing fewer than 57 correct binary responses out of 96 (the one-tailed  $p < 0.05$  cutoff under coin-flip guessing), and “ceiling” is defined as 88 or more correct (i.e., the same distance from 100% as floor is from 50%), then of the 60 participants in the trimmed sample, only 9 were still at floor, and 7 at ceiling. Of the 30 participants excluded for low performance, all but 2 were at floor, and of the 30 excluded for high performance, all but 4 were at ceiling.

## Full Sample Analysis

Both the binary and continuous responses were analyzed, yielding qualitatively similar results. In the interest of concision, we report only the latter here. Mean scores were computed for each participant at each block and entered into an ANOVA with between-subjects factor Context Condition (five levels: Uniform, Smooth (High Variance), Smooth (Low Variance), Repetition (High Variance) and Repetition (Low Variance)), and within-subjects factor Block (1 through 4). Four planned contrasts were used for the Context factor: the first three concerned the Repetition and Smooth groups, corresponding to main effects of (1) distribution type and (2) variance, and (3) the interaction between distribution and variance; the last comparison contrasted the two Repetition groups with the Uniform group.

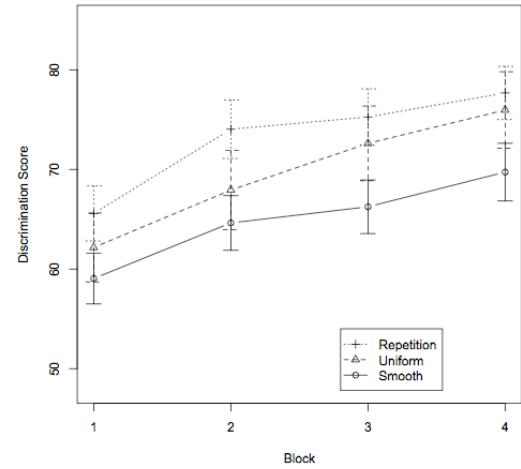


Figure 3: Mean Discrimination Scores by Context Condition and Block, Full Sample

The main effect of Block was significant ( $F(3, 345) = 30.59, p < 10^{-15}$ ) but the Block X Context interaction was nonsignificant ( $F(12, 345) = 0.55, \text{n.s.}$ ). Of the contrasts among context conditions, only the contrast between the Repetition and Smooth groups reached significance ( $F(1, 115) = 5.63, p < 0.02$ ). The contrast between the Repetition group and the Uniform group was nonsignificant ( $F(1, 115) = 0.03, \text{n.s.}$ ), as were the contrast between the High and Low Variance groups ( $F(1, 115) = 1.08, \text{n.s.}$ ) and the Distribution X Variance interaction ( $F(1, 115) = 0.12, \text{n.s.}$ ). Means and standard errors for each block and each group (collapsing the High and Low Variance groups) are displayed in Fig. 3.

## Trimmed-Sample Analysis

The above analysis was repeated using only those participants in the second and third quartiles within each context group, as determined by total number correct collapsed across blocks. The effect of Block was significant ( $F(3, 165) = 26.89, p < 10^{-13}$ ), but the Block X Context interaction was not ( $F(12, 165) = 1.12, \text{n.s.}$ ). Of the contrasts among context conditions, only the contrast between the Repetition group and the Smooth group reached significance ( $F(1, 55) = 16.16, p < 0.001$ ). The contrast between the Repetition group and the Uniform group was nonsignificant ( $F(1, 55) = 0.41, \text{n.s.}$ ), as were the contrast between the High and Low Variance groups ( $F(1, 55) = 1.58, \text{n.s.}$ ) and the Distribution X Variance interaction ( $F(1, 55) = 0.004, \text{n.s.}$ ). Means and standard errors for this trimmed sample are displayed in Fig. 4.

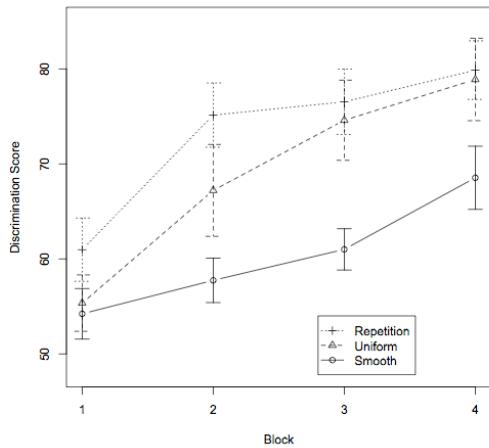


Figure 4: Mean Discrimination Scores, 2<sup>nd</sup> and 3<sup>rd</sup> Quartiles Only

## Discussion

The present experiment set out to investigate the contribution of rational, generative “explaining away” to induction of an abstract repetition rule over a set of tone sequences. In the context of sequences of musical tones, repetition has a dual nature, first as an identity relation between two consecutive events, and second as an interval of magnitude zero between two tones on a continuum. Hence when a repetition occurs, it is ambiguous which of these two descriptions should be attached to it. The central finding was that adult humans appear to take into account a global “smoothness constraint” on melodies, which have a statistical tendency to move in small intervals, to set a baseline expectation for the rate of repetitions. This reduces the informational value of a repeated tone as a cue to an abstract rule.

The secondary prediction was that participants in the Repetition condition (in which repetitions are uniquely frequent and cannot be explained except by representing them explicitly) would be inclined to downplay the ordinal relations among tones in their representation of the environment, focusing instead on abstract, “discrete” relations like “same” and “different”. Since these are precisely the relations needed to learn the repetition grammar, participants in this group were expected to learn the rule more easily than those in the Uniform group. Support for this prediction is tenuous at best: although scores in the Repetition group were numerically higher than those in the Uniform group, this difference failed to reach significance. Exploratory analyses suggested that an advantage for the Repetition group may be present in the early blocks, disappearing later, but the difference was only marginally significant, and in any case the analysis was *post hoc*. It may be that a larger sample is needed to detect a difference if one indeed exists.

At first glance, the presence of differences in participants’ use of information resembles situations in which learners come to focus on features that are predictive of a relevant task outcome, filtering out redundant information (Haider and French, 1996; Pellegrino, Doane, Fischer and Alderton, 1991; Doane, Sohn and Schrieber, 1999). Although many of the same mechanisms may be involved here, the nature of the learning is somewhat different. Whereas in the preceding experiments participants were engaged in a specific task all along, in the present experiment the key manipulation occurs before participants become aware of what it is they will be asked to learn. As such, it is not simply a matter of repetition being predictive of a particular response (or even of other stimulus features); rather this result suggests that learners in this experiment are creating an *explanatory* model of the alien environment, and forming hypotheses about how their input is being generated.

Although ultimately the value of explanation may be connected to the future ability to make predictions, the absence of explicit behavioral demands frees learners to pursue a general goal of understanding the underlying nature of the environment. Here, in the Smooth environment repetitions do not appear to be an essential component of the environment at all, whereas in the Repetition environment it is necessary to represent them in order to understand the distribution of intervals. This concept of the learning process as rational hypothesis testing fits nicely into the wealth of recent literature using Bayesian models to capture aspects of cognitive functioning (see, e.g., Tenenbaum, Griffiths and Kemp (2006), for a review).

The present set of findings is of great relevance to the rule-learning literature initiated by Marcus, et al. (1999), and is particularly supportive of the conjecture by Dawson and Gerken (2009) that 7.5-month-olds may have “learned to fail” at learning AAB rules due to the acquisition of knowledge about tonality and the smoothness of natural melodies. We are currently carrying out a version of the present experiment adapted to infants to determine whether the explaining away process observed here in adults comes into play in infancy as well. If so, it will add a new explanatory tool to be applied to the puzzle of why formally analogous rules are easier to learn in some contexts than others. More generally, the sort of “metalearning” observed here may play an important role in the formation of apparently domain-specific biases and constraints. In general, when a potential role for differential experience exists, caution should be exercised before proposing innate biases.

Finally, in order to explain away, learners must be explaining in the first place. The present findings add to a growing body of evidence (Gopnik, 1998; Schulz and Bonawitz, 2007; Xu and Garcia, 2008; Gerken, 2010) that learning is a lot like science: in addition to making specific predictions, an important role of cognition is to build explanatory models of the environment, and to construct and test hypotheses about why the world works the way it does.

## Acknowledgments

This research was supported by an NSF Graduate Research Fellowship to Colin Dawson, as well as NIH grant R01 HD042170 to LouAnn Gerken. The authors also wish to thank Brianna McMillan and Kailey Tucker of the Tweety Language Development Laboratory at the University of Arizona for their assistance with data collection.

## References

Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217-243.

Ciaramita, M., & Johnson, M. (2000). Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. In Proceedings of the 18th International Conference on Computational Linguistics, 187-193, Saarbrücken, Germany.

Dawson, C. (2007). Infants Learn to Attend to Different Relations When Forming Generalizations in Different Domains. Unpublished Master's thesis.

Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-Month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, 111(3), 378-382.

Doane, S., Sohn, Y. W., & Schrieber, B. (1999). The role of processing strategies in the acquisition and transfer of a cognitive skill. *Journal of Experimental Psychology: Human Perception and Performance*, 25(5), 1390-1410.

Eerola, T., & Toivainen, P. (2004). MIDI Toolbox: MATLAB Tools for Music Research. University of Jyväskylä: Kopijyvä, Jyväskylä, Finland.

Folstein, J. R., Van Petten, C., Rose, S. A. (2007) Novelty and conflict in the Categorization of complex stimuli. *Psychophysiology*, 45, 467-479.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, 7, 287-292.

Gerken, L. A. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, 115(2), 362-6.

Gerken, L., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms. *Language Learning and Development*, 4(3), 228-248.

Gopnik, A. (1998) Explanation as orgasm. *Minds and Machines*, 8(1), 101-118.

Hannon, E. E., & Trehub, S. E. (2005). Tuning into musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, 102(35), 12639-12643.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30(3), 304-337.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N. Z., Marcus, G. F., Rabagliati, H., & Slemmer, J. A. (in press). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy*.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review Psychology*, 55, 271-304.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception and Psychophysics*, 52(6), 599-608.

Marcus, G., Fernandes, K., & Johnson, S. (2007). Infant Rule-Learning Facilitated by Speech. *Psychological Science*, 18, 387-391.

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

Maye, J., Werker, J., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, 101-111.

Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, 319, 1849-1851.

Pellegrino, J. W., Doane, S. M., Fischer, S. C., & Alderton, D. (1991). Stimulus complexity effects in visual comparisons: The effects of practice and learning context. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 781-791.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110-114.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, 37, 74-85.

Saffran, J. R., Pollack, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, 105, 669-680.

Schulz, L., & Bonawitz, E.B. (2007) Serious fun: Preschoolers play more when evidence is confounded. *Developmental Psychology*, 43(4), 145-150.

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science: A Multidisciplinary Journal*, 32(2), 418-444.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7), 309-318.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13), 5012-5015.