DOMAIN-SENSITIVE TUNING OF RELATIONAL GENERALIZATION
IN THE FIRST YEAR OF LIFE

by

Colin Dawson

A Thesis Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

STATEMENT BY AUTHOR

SIGNED: Colin Dawson

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

____________________________	10/24/2007
LouAnn Gerken	Date
Professor of Psychology	

TABLE OF CONTENTS

ABSTRACT

Two age groups of infants were tested for their ability to learn an AAB or ABA repetition generalization in sequences of musical chords. The 4-month-olds, but not the 7.5-month-olds, successfully learned the generalization. Another group of 7.5-month-old infants successfully learned a generalization across melodies that all ended on a particular scale degree, even though the key of the melodies was varied. A survey of a musical corpus of children's songs reveals that AAB and ABA patterns do not occur more frequently than chance, while phrases frequently end on particular scale degrees. Together, these findings suggest that infants learn to constrain the set of generalizations they consider in order to favor those that rely upon features of the input that have proved reliable in their previous experience, specifically experience with a particular input domain. This raises the possibility that experience may play a significant role in parsing infants' environments into domains.

INTRODUCTION

Statistical Learning

A great deal of research in language acquisition, as well as in human and animal learning more generally, has focused on the ability of infants and others to extract statistical structure from their input. In a now classic set of experiments, Saffran, Aslin and Newport (1996) and Aslin, Saffran and Newport (1998) showed that human infants and adults could segment streams of syllables into constituent pseudowords by tracking transitional probabilities between successive syllables. In these studies, subjects are familiarized with an uninterrupted stream of syllables constructed by stringing together three-syllable pseudowords from an artificial language. Subjects are then presented with trisyllables in isolation, some of which are the constituent pseudowords from familiarization, others of which are "part words" consisting of the end of one word followed by the beginning of another. Even when words and part words are frequency matched in familiarization, both infants and adults successfully distinguish them, presumably on the basis of the transitional probabilities from one syllable to the next.

Infants can also use featural distributional information to form phonetic categories (Maye and Gerken, 2000; Maye, Werker and Gerken, 2002). Here, infants hear several consonant-vowel (CV) syllable tokens, where the voice onset time (VOT) of the initial consonant is varied along a continuum. When the VOT is distributed unimodally, infants act as though there is only one phonetic category. When, on the other hand, VOT values form a bimodal distribution, infants act as though two categories are present.

In addition to extracting statistics about featural properties of their input, infants can also discover generalizations based on more abstract, relational information. Marcus, Vijayan, Bandi Rao and Vishton (1999) showed that infants could learn an artificial grammar where “sentences” possessed an AAB or an ABA structure over syllables. That is, although the slots in the grammar could be filled by any syllable, the grammar dictated that the syllables in the two “A” positions be identical. After being familiarized for two minutes on three tokens each of sixteen “sentences”, 7.5-month-old infants successfully discriminated novel sentences, containing entirely novel syllables, that were consistent with their training grammar, from those that possessed the opposite pattern. Gómez and Gerken (1999) showed that 12-month-old infants trained on syllable sequences generated by a finite-state grammar appear to learn the underlying first-order dependencies, discriminating novel strings that could be generated by the grammar from those that could not. Furthermore, they showed that infants could generalize these dependencies to a novel vocabulary of syllables, apparently extracting the abstract graph shape of the finite-state grammar, including the presence of loops, while allowing the nodes in the graph to be filled in by new syllables. This ability contains the result of Marcus, et al. (1999) as a special case, though the infants in the Gómez and Gerken (1999) study are four months older.

Constraints on Learning

While powerful learning mechanisms are clearly present, it is less clear how learning is constrained. Constraints on learning are necessary for two primary reasons.

First, real cognitive systems have limited memory and processing resources. In contrast, the number of possible sources of statistical structure in the environment is unbounded. Therefore, any real learning mechanism must have some way to choose which sources of statistical information to attend to. The second problem faced by an unconstrained learning mechanism is that of overgeneralization. Nowhere is this problem more apparent than in language acquisition, by virtue of the fact that not only must the correct structure be perceived, but new utterances must be generated that contain the proper structure. Whereas undergeneralization (that is, forming a hypothesis about the world's structure that is too restrictive) and misgeneralization (forming a hypothesis that partially overlaps with the true structure) can be corrected upon encountering positive evidence that is inconsistent with one's mental model, recovering from a hypothesis that predicts a superset of the real data would seem to require negative evidence in order to be corrected. It has long been observed that children do not appear to receive overt negative evidence (e.g. Marcus, 1993), and when they do it appears to be largely ignored (Brown and Hanlon, 1970). These observations have been taken as proof that language learning would be impossible without strong constraints on learning mechanisms.

The problem of selecting among an effectively infinite number of possible sources of information has had several classes of (non-mutually exclusive) proposed solutions. First, it has been proposed that an innate knowledge base, built up over evolutionary time and encoded in the genome, guides learners to seek out just the right kinds of structure. In the domain of language acquisition, this knowledge could take the form of Universal Grammar, which would contain a set of principles, common to all languages, and a set of parameters, each of which can take on one of a small finite number of settings, depending on what language is being spoken (Chomsky, 1981). On this view, learning would come into play only in setting these parameters, as well as forming associations between signifiers and signifieds. Another, more general, potential class of constraint is a tendency to prefer "simpler" structural models. While it is difficult to define simplicity objectively, several recent accounts of constraints on language learning have this flavor. Gerken (2006), on the basis of data showing that infants only make an abstract, relational generalization when a more concrete, token-based generalization is not available, has suggested that infants' generalizations are consistent with the predictions of a Bayesian hypothesis-selection account, preferring hypotheses that assign a higher probability (likelihood) to the observed data, all else being equal. Similarly, Gómez (2002) has shown that both adults and infants do not learn non-adjacent sequential dependencies unless the adjacent statistics are sufficiently unreliable. These results might be accounted for by a version of Occam's Razor that gives preference to hypotheses that restrict the range of data that they can generate without relying on ad hoc fine-tuning of parameters.

A second class of constraint takes the form of limits on perceptual systems. While the world might in fact contain an effectively infinite number of information sources, only a relatively small subset of these sources may be detectable by human perception (e.g. Saffran, 2002, 2003; Newport and Aslin, 2004). For example, human vision is sensitive only to a small range of wavelengths in the electromagnetic spectrum.

Any structure of frequencies outside this range cannot in principle be considered, since the signal is not transduced by the eye.

A third way in which learning may be constrained is through preference for correlated cues (e.g. Morgan and Demuth, 1996; Gerken, Wilson and Lewis, 2005). These studies show that when multiple cues to a structure co-occur, infants more readily learn that structure than if only one cue is present.

The final proposal, and that which I will focus on here, is that learners' prior experience guides future learning. There are at least two ways in which this might play out. First, learners might develop more compact and/or sophisticated representations of their input, which can serve as a scaffold for learning successively more complex patterns (e.g. Saffran and Thiessen, 2003; Lany and Gomez, 2004; Lany, Gomez and Gerken, 2007). Second, learners may develop attentional weighting schemes, preferring to consider sources of information that have proved reliable in the past (Kruschke, 1992; 2003; 2005; Tomlinson and Love, 2006).

The Role of Domain

The information selection problem is complicated by the fact that cues that are reliable in one scenario are often unreliable in another. For instance, while color is critically important in determining the ripeness of fruit, it plays little role in extracting the meaning of a passage of text. On the other hand, while an object's orientation in space can be critical in distinguishing among printed letters (e.g. distinguishing "p" from "b" from "q"), spatial orientation is probably irrelevant in distinguishing fresh from rotten apples. Hence, a learner cannot simply use a single weighting scheme, privileging some sources of information and ignoring others, in all situations. However, similar kinds of input are likely to have similar kinds of structure, and so if a learner is able to classify input as belonging to a particular domain, then she can potentially bring to bear knowledge about that domain in choosing which information to attend to and which to ignore. In essence, stability of the set of relevant information across scenarios can be said to be the defining feature of a domain.

An important question is to what extent differentiation of domains is the purview of learning, and to what extent innate constraints guide learners to employ particular learning strategies when faced with particular kinds of input. It has been proposed that language constitutes a domain to which specialized learning mechanisms are devoted (Chomsky, 1980). However, there is mounting evidence that at least some of the learning mechanisms that apply to language are not dedicated to language. For one, humans appear in some cases to learn analogous patterns equally well in linguistic and non-linguistic input. Saffran, Johnson, Aslin and Newport (1999) have shown that both adults and infants extract constituent sequences from longer sequences of tones by using transition probabilities. This result replicates the earlier results of Saffran, Aslin and Newport (1996) and Aslin, Saffran and Newport (1998) showing that such segmentation is possible with sequences of syllables. Similarly, Fiser and Aslin (2001; 2002) have shown that infants and adults can use conditional co-occurrence probabilities to segment visual scenes into constituent groups of shapes. Another set of evidence shows that non-human animals succeed at some of the same statistical learning tasks as humans. Hauser, Newport and Aslin (2001) and Toro and Trobalon (2005) show that cotton top tamarins and rats, respectively, succeed at segmenting sequences of syllables using transitional probabilities. As these animals are not evolved to use language, their success in these tasks cannot be attributed to a language-specific learning mechanism.

While great generality is exhibited in statistical learning at the level of segmentation, it is less clear to what extent the more abstract, relational learning of the sort exhibited by the infants in Gómez and Gerken (1999) and Marcus, et al. (1999) is applicable to non-linguistic domains. Indeed, Marcus, Johnson and Fernandes (2007) have shown that 7.5-month-old infants fail to learn an AAB/ABA generalization over tone sequences where they succeed with precisely analogous sequences of syllables. Their conclusion is that learning this kind of generalization requires a variable-binding ability that is evolved in humans specifically for dealing with language. However, failure to learn in one domain what can be learned in another need not be attributed to an innate distinction between learning mechanisms as they apply to different kinds of input. Complete domain-generality and innate domain-specificity are not the only possibilities for learning mechanisms. A third possibility is that such learning abilities begin with the *potential* to apply to any kind of input, but come to function differently in each domain as experience reveals the presence of different structures and predictive relationships.

Adaptation of Learning and Perception

Within domains, the narrowing and specialization of learning abilities to adapt to the environment is a well-documented phenomenon. In language, the classic study of

Werker and Tees (1984) showed that, during the first year of life, infants lose sensitivity to phonetic contrasts that are not used phonologically in their native language. At the same time, infants' sensitivity to native phonological contrasts can actually be enhanced by exposure to statistical distributions that suggest the presence of two categories (Maye and Weiss, 2003). Event-Related Potential (ERP) data from Rivera-Gaxiola, Silva-Pereyra and Kuhl (2005) supports both of these findings, showing both a decrease in ERP response to non-native contrasts and an increase in response to native contrasts between 7 and 11 months of age. Yet, the ERP response to nonnative contrasts does not disappear at 11 months, but instead seems to shift to a different signature, with two distinct subgroups of infants displaying different ERP signatures.

Interestingly, even adults show sensitivity to variations that are not used contrastively in their native language, provided those variations constitute a valid cue to lexical identity. Gaskell and Marslen-Wilson (1996) showed that phonetically altering a segment to an allophone that could not be produced in the context in question blocked priming of the word in a cross-modal priming paradigm, indicating that the altered allophone was treated as distinct from the original allophone. Further evidence that subphonemic information serves an important purpose in ecologically meaningful linguistic processing comes from Johnson and Jusczyk (2001), who showed that 8-month-old infants use coarticulatory information to segment a speech stream, even when this phonetic information conflicted with the transitional statistics. They found a similar result when stress cues were pitted against transitional statistics: namely that 8-month-olds preferred to treat trochaic feet as words even when the statistics indicated otherwise.

However, Thiessen and Saffran (2003) found that slightly younger infants (7 months old) segmented according to statistics and *not* stress when the two conflicted, suggesting that infants may initially use statistics to segment, but subsequently discover that the constituent words revealed by statistics have a characteristic stress pattern, and so later come to use stress as a cue to constituency. That the use of stress in segmentation appears later than the use of statistics stands to reason, since if stress-based regularities are to be discovered, an initial segmentation must first be performed.

An instance of a narrowing of learning in higher-level generalization tasks comes from Gerken and Bollt (in press), who show that, while 7.5-month-olds are equally capable of learning both a "natural" stress rule (i.e. one found in languages of the world) and an "unnatural" rule (one that does not exist in any known language), by 9 months of age, infants can learn only the natural rule. They familiarized infants with three- and five-syllable nonce words that followed one of two stress rules. When the rule was "stress heavy syllables" (the natural rule), both the 7.5-month-olds and the 9-month-olds discriminated grammatical from ungrammatical test items. When the rule was "stress syllables beginning with /t/, however, the 7.5-month-olds but not the 9-month-olds discriminated the test items. The success of the 7.5-month-olds in learning the unnatural rule is evidence against the hypothesis that infants are born with knowledge about which cues play a role in determining stress in the world's languages. Rather, the developmental narrowing suggests that the English-exposed infants in the study are picking up on a correlation (albeit imperfect) between syllable coda information and stress in English, and hence learn that heaviness is a useful cue to stress. On the other hand, the lack of correlation between onset information and stress appears to lead the 9-month-olds to ignore onset when learning a prosodic generalization.

Infants in a study by Cristia and Seidl (in press) show a narrowing in their phonological generalization abilities between 7 and 14 months of age. In this study, infants were familiarized with sequences that exhibit particular phonotactic regularities. In one case, the pattern depends on a phonological class consisting of nasals and stops. In traditional feature-based phonologies, the feature [- continuant] is both necessary and sufficient to characterize nasals and stops, meaning that these segments constitute a natural class. The 7-month-olds successfully generalized the phonotactic regularity to novel stops. By 14 months, however, the generalization was no longer made. Cristia and Seidl suggest that this loss in generalization is due to the absence of phonological rules in English that hinge on the [- continuant] natural class. It would appear that English-exposed infants learn to ignore the phonological dimension(s) necessary to pick out the class of [- continuant] segments in learning phonological generalizations, due to its lack of reliability in their input.

There is evidence for the specialization of learning in the domain of music as well. Studies by Saffran and Griepentrog (2001) and Saffran (2003) show that when absolute pitch cues and relative pitch cues are put into competition, infants segment tone streams according to absolute cues, while adults rely on relative cues. This change would seem to be adaptive to a musical environment where melodies can be transposed freely and retain their identity. Interestingly, when absolute pitch cues are rendered uninformative, infants will segment according to relative pitch cues (Saffran, Reeck, Niebuhr and Wilson, 2005).

Hannon and Trehub (2005) tested North American adults, Bulgarian and Macedonian adults, and North American infants on a task requiring them to rate metrically altered folk melodies for their similarity to the original. For melodies with a simple temporal ratio duration structure, all groups rated altered melodies that preserved the global metrical structure as more similar to the original than melodies altered to violate the metrical structure. However, when melodies with complex meters were used, the North American adults showed no difference in their ratings of structure-preserving versus structure-violating alterations. Both the infants and the Bulgarian and Macedonian adults appeared to perceive the structure-preserving melodies as more similar to the original. This result is somewhat analogous to studies showing a loss of phonetic discrimination for nonnative contrasts, as the adults whose native culture contained music with primarily simple ratio meters appear to have lost the ability to discriminate between structure-preserving alterations and structure-violating alterations, whereas both infants and adults from a culture that regularly employs complex meters discriminated well.

Trainor and Trehub (1992) tested infants' and adults' abilities to detect small changes in melodies. In one condition a single note was raised or lowered by 4 semitones, i.e., 1/3 of an octave, but remained within the diatonic scale of the original key. In the other condition, a note was raised or lowered by a single semitone, such that the resulting pitch was not in the diatonic scale of the original key. While infants were equally able to detect these two changes, adults much more easily detected the change that violated the scale structure of the melody, despite the larger acoustic change in the scale-preserving alteration. This finding suggests that adults learn a schema, corresponding to global tonality, which produces perceptual assimilation effects for melodic changes that do not alter tonality.

Taken together, the above findings seem to converge in support of the idea that, as learners gain experience with different parts of their environment, they bring to bear knowledge about the relevance and reliability of various environmental sources of information as they continue to learn. Such an account appears to be at odds with theories that assume a primary or even exclusive role of a rich base of innate knowledge in constraining learning.

The goal of the present set of experiments is to explore the possibility that infants reorganize their learning apparatus in a way that not only focuses attention on cues that have proved reliable in the past, but that they do so differently, and simultaneously, for different domains. The experimental procedure is based on the task used by Marcus, et al. (1999), and again by Marcus, et al. (2007). Infants are familiarized with a set of sequences, all of which share a particular set of relations. They are then tested on novel sequences, half of which are consistent with the generalization from familiarization, the other half of which adhere to a different generalization. As is typical in infant research, a consistent preference for either the consistent or inconsistent stimuli is taken to reflect discrimination, and therefore learning of the pattern during familiarization.

EXPERIMENTS

Experiment 1

Methods

Participants

Eighteen infants between the ages of 3.5 months and 4.5 months old (mean age 19 weeks) were recruited from the Tucson area. Seven were female. Data from six additional infants was collected, but was excluded from analysis due to these infants' failure to complete the minimum required number of test trials. In order to be eligible, infants had to be at least 37 weeks to term, weigh at least 5 lbs 8 oz at birth, and have no history of speech or language problems in their immediate (nuclear) family.

Materials

Three-note triads were built on each of the twelve chromatic tones between middle C and the B above middle C. Half were pseudorandomly selected to be major triads, while the other half were minor triads, with the constraint that no note occurred in the same position (root, third or fifth) in any two chords. Eight of these twelve (four major and four minor) were pseudorandomly assigned to the familiarization phase, the remaining four to the test phase (assignment was pseudorandom because of the major/minor constraint). The sets of chords for each phase (familiarization and test) were further pseudorandomly divided in half into an A group and a B group, again preserving the equality between major and minor and with two additional constraints. No chord in the A group shared more than one tone with any chord in the corresponding B group, and in half of the possible pairings of an A chord with a B chord, the root of the former was a higher note, in the other half it was lower.

The familiarization phase contained four different chords in each of the A and B groups; the test phase two in each. Three-chord phrases were created for each of two "grammars" – the AAB grammar and the ABA grammar. Phrases in the AAB grammar consisted of an A element repeated twice, followed by a B element, while those in the ABA grammar consisted of an A element followed by a B element followed by the first element again. Every possible combination of A element and B element was represented, for a total of sixteen phrases in each grammar in the familiarization phase and four phrases in each grammar in the test phase. Each phrase was 2500 ms – 625 ms for each chord and 625 ms of silence at the end. No two of the twenty total phrases contained exactly the same relative pitch patterns.

Two familiarization trials were constructed, one using the AAB grammar, the other using the ABA grammar. Each trial consisted of three blocks, where a block contained each of the sixteen phrases in the grammar once in random order. The three blocks each had a different random order of the phrases, but the same random orders were used for the AAB trial and the ABA trial – that is, if phrase $A_1A_1B_3$ occurred in the first position in the AAB trial, then the phrase $A_1B_3A_1$ occurred in the first position in the ABA trial, and so on. There were no breaks either between phrases in a block or between blocks (though each phrase contained 625 ms of silence at the end). Each trial was a total of 2 minutes in length.

Test trials were constructed in the same way, except that two trials were constructed for each grammar. Again, each trial consisted of three blocks, and each block contained each of the four phrases in the grammar. Again the order of trials within each block was random, but the order was matched within each pair of test trials (one AAB and one ABA trial). Each test trial was 30 seconds in length.

Procedure

The headturn preference procedure (Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 2005) was used. Infants were seated on a parent's lap in a small room. The parent listened to popular music through headphones, in order to mask the music heard by the infants and thereby prevent inadvertent influence on the infant by the parent. Parents were instructed that they could end the study at any time by signaling to the observer via the camera. During the familiarization phase, a light directly in front of the infant flashed until the observer judged the infant to be looking at it, at which point a light on the left or right would begin flashing. When the infant looked first at the side light and then away for two consecutive seconds, the center light would flash again, and the cycle would begin again. This continued for the duration of the familiarization music, which played uninterrupted to its conclusion. In this stage there was no correspondence between infants' looking behavior and the music.

After the familiarization sequence ended, the test phase began immediately. The flashing lights behaved the same way, except that now the sound was contingent on the infant orienting to a side light. Each time a side light began flashing and the infant oriented toward it, one of the four test trials would play, continuing until either the infant looked away for two consecutive seconds or the test trial reached its conclusion. Each infant heard three blocks of test trials. Each of the four trials occurred once per block, the order randomized separately for each infant.

The observer controlling the presentation of the music sequences could not hear anything happening within the booth, and so was blind to the category of each test trial. The computer automatically recorded the looking time for each test trial.

Inclusion Criteria

Since the AAB and the ABA grammars cannot be distinguished by the first element alone, test trials that were terminated by a look-away before the second chord of the first phrase was played conveyed no grammatical information to the infants. As such, trials terminated before 2 seconds had elapsed were excluded from analysis. All data from any infant who did not have sufficiently long looking times to at least three trials from each grammar was excluded from analysis. Data from six infants was excluded for this reason.

Results

Average looking times to each of the two grammars were computed for each infant. For purposes of analysis, the experiment was treated as having a single independent variable, grammaticality, with two levels, consistent and inconsistent. A related-samples t-test was performed on the pairs of looking times, revealing a significant effect of grammaticality, $t(17) = 3.19$, $p(\text{2-tailed})<0.005$, with infants looking longer during inconsistent trials, $M_{consistent}=12.28$ sec, $M_{inconsistent}=15.07$ sec; $SE_{difference} = 1.05$ sec; Cohen's $d = 0.64$.

Discussion

Infants' consistent preference for the novel grammar indicates that they are able to learn the abstract relational generalization that defines the grammar to which they are exposed. This result with non-linguistic stimuli contradicts the claim by Marcus, et al. (2007) that infants' ability to learn such a generalization relies on an algebraic, symbol-manipulation mechanism that is domain-specific to language. The results are consistent with the notion that the infants in Marcus, et al. (2007) fail to discriminate the two grammars due to their learned expectation that sequential dependencies of the sort that define an AAB or ABA grammar are not among the set of characteristics that are meaningful in the musical domain.

Since the stimuli used in Experiment 1 were not identical to those used by Marcus, et al. (2007), the possibility remains that the discrepancy in generalization abilities between the infants in the two studies is not due to the difference in age, but rather to some difference in the nature of the stimuli. Experiment 2 is intended to address this possibility.

Experiment 2

If the ability of the 4-month-old infants in Experiment 1 to learn the abstract generalization required to discriminate the test stimuli is indeed due to their relative inexperience with music, while the failure of the 7.5-month-olds in the Marcus, et al. (2007) study to discriminate the two grammars is due to learned expectations about the structure of music, it should be expected that 7.5-month-olds tested with the stimuli used in Experiment 1 should fail to discriminate the two grammars. Experiment 2 serves as an attempt to replicate the null finding of Marcus, et al. (2007) using the stimuli from Experiment 1.

Methods

Participants

Eighteen infants between the ages of 7 and 8 months (mean age 32 weeks) were recruited from the Tucson area. Eight were female. Data from six additional infants was excluded from analysis due to these infants' failure to complete the minimum number of test trials. A priori eligibility requirements were identical to those in Experiment 1.

Materials and Procedures

Stimuli and procedures were identical to those used in Experiment 1.

Inclusion Criteria

Minimum looking time requirements were identical to those in Experiment 1. Six infants' data was excluded due to an insufficient number of trials meeting minimum looking time criteria.

Results

Average looking times for each of the two levels of the independent variable (consistent and inconsistent) were computed for each infant. A related-samples t-test revealed no effect of grammaticality, $t(17) = 0.18$, $p > 0.75$, $M_{consistent} = 9.52$ sec, $M_{inconsistent} = 9.17$ sec, $SE_{difference} = 1.04$ sec; Cohen's $d = 0.08$).

The looking times from Experiments 1 and 2 were entered together into an ANOVA with between subjects factor age (4 months and 7.5 months) and within subjects factor grammaticality (consistent and inconsistent). The effect of age was significant, $F(1,34)=6.01$, $p<0.05$, with 4-month-olds looking longer overall ($M_{4m}=13.68$ sec, $M_{7.5m}=9.35$ sec). Most importantly, there was a significant age X test type interaction, $F(1,34)=4.35$, $p<0.05$.

Discussion

The 7.5-month-olds' failure to discriminate consistent from inconsistent test trials serves to replicate the finding of Marcus, et al. (2007) that infants at this age are not able to learn the abstract generalization about sequential dependencies when the stimuli are musical in nature. Coupled with the finding from Experiment 1, the present results support the notion that a domain-sensitive informational filter is developed between 4 and 8 months of age that excludes the sort of sequential dependency defining the musical grammars from the set of patterns that are expected in the musical domain. It is this experience-dependent filter, and not an innate, domain-specific symbol manipulation mechanism that accounts for the difference in findings between linguistic and musical stimuli in Marcus, et al. (2007).

While the first two experiments convincingly demonstrate that 4-month-olds and 7.5-month-olds learn about music differently, they do not necessarily show that it is an informational selection process, rather than a broad loss of ability to learn about relational patterns in music, that accounts for the 7.5-month-olds' failure to learn the generalization required in Experiment 2. Experiment 3 explores further the nature of 7.5-month-olds' abilities to learn about music.

Experiment 3

Although sequential dependencies among individual chords are not characteristically found in the music infants typically hear, such music is not entirely devoid of relational structure. Durational ratios (rhythm and meter), as well as frequency relationships (tonality) convey important structural information in music. Rhythm and meter are central in determining what sort of dancing, if any, goes along with music; tonality can often evoke emotional responses (in the western European tradition, a major key is stereotypically perceived as “happier” than a minor key); tonality-dependent chord progressions can convey a sense of anticipation or finality at a phrase’s end.

If indeed infants at 7.5 months of age have learned enough about the structure of their culture’s music to have a sense of what kinds of patterns to expect and not to expect, this knowledge should manifest itself not only in an inability to learn atypical patterns, but also in an ability to learn musically relevant patterns. Experiment 3 asks whether infants at this age can learn a generalization about the relationship of a phrase-final chord to the tonal structure of the phrase. In this experiment, infants are familiarized with a set of melodies, all of which end on a chord with the same relationship to the key. In one condition, all of the melodies end on a I (tonic) chord, and in the other they all end on a V (dominant) chord. As in experiments 1 and 2, infants then hear test trials, half of which adhere to the “grammar” with which they were familiarized, while the other half adhere to the opposite grammar.

In typical folk music, not least children’s music, the last phrase in a larger phrase ends on the tonic, while the second-to-last phrase often ends on the dominant. Hence, hearing a phrase that ends on a dominant chord leads a listener to expect more, while hearing a phrase ending on a tonic chord conveys a sense of completion (Piston, 1969).

Methods

Participants

Eighteen infants between the ages of 7 months and 8 months were recruited from the Tucson area. Ten were female. Data from seven additional infants was excluded from analysis due to these infants' failure to meet a looking time criterion on a minimum number of test trials. A priori eligibility requirements were identical to those in experiments 1 and 2.

Materials

Twelve distinct four-note "carrier" melodies were composed, all in a major key. Each melody was then transposed, so that no two melodies were in the same key. Four melodies were arbitrarily chosen to be test melodies; the remaining eight were used as familiarization melodies. The "ends in 1 (e1)" grammar was created by appending a tonic chord to the end of each carrier melody. The "ends in 5 (e5)" melody was created by appending a dominant chord to the end of each carrier melody. In addition, the chord sequence I-V-I was prepended to each melody (in the appropriate key). This was done to make more apparent what key the short melodies were in. Every melody had the same rhythmic pattern: the initial I-V-I sequence comprised three quarter notes followed by a quarter rest, the four carrier tones were each an eighth note, and the final chord was a quarter note. A quarter rest followed each melody. Overall, the melodies were two measures long, in a 4/4 (common time) meter, at a tempo of 120 beats per minute. Each melody was therefore four seconds long, including the final rest.

Familiarization trials contained four blocks, where each block contained one instance of each melody in the grammar, in a different random order for each block. The order of the carrier phrases was identical between the e1 familiarization sequence and the e5 sequence. Each familiarization trial was 2 minutes and 8 seconds long.

Test trials contained two blocks, where again each block contained one instance of each of the four melodies in the grammar, in a different random order for each block. Two test trials were created for each grammar, with different orders of the melodies. Each e1 test trial contained the same order of carrier phrases as one of the e5 trials. Each test trial was 32 seconds long when played in its entirety.

Procedures

The procedures for Experiment 3 were identical to the procedures used in Experiments 1 and 2.

Inclusion Criteria

Since the information distinguishing the e1 grammar from the e5 grammar did not occur until the end of a phrase, test trials that were terminated by a look-away prior to the final chord of the first phrase in the trial could not be identified as one grammar or the other. For this reason, test trials with looking times less than three seconds were excluded from analysis. All data from any infant who did not have sufficiently long looking times to at least three trials from each grammar was excluded from analysis. Data from six infants was excluded for this reason.

Results

As in Experiments 1 and 2, Experiment 3 was treated as having a single independent variable (grammaticality) with two levels (consistent and inconsistent).

Average looking times were computed for each of the two test trial types, for each infant. A related-samples t-test performed on these times revealed a significant effect of grammaticality, with infants looking longer to inconsistent test trials ($t(17) = 2.11$, $p < 0.05$, $M_{consistent} = 8.60$ sec, $M_{inconsistent} = 10.66$ sec, $SE_{difference} = 0.98$ sec; Cohen's $d = 0.50$).

Discussion

The 7.5-month-olds' success in learning the generalization in experiment 3 rules out the possibility that infants at this age have simply lost the ability to learn abstract relations in musical stimuli. This result further strengthens the hypothesis that, as infants gain experience in the musical domain, they develop a domain-specific "attentional mode" which, in the domain of music, shifts resources away from the kind of sequential dependency information necessary to learn the generalization in experiments 1 and 2, instead allocating resources to information about tonality, including information about the role of individual chords within the broader key system.

It should be noted that it is not clear whether 4-month-olds would succeed at this task. Given the evidence suggesting that younger infants rely more heavily on absolute pitch cues rather than relative pitch cues, it may be that infants at 4 months do not have a sufficiently well-developed understanding of global structure in music to pick up on a generalization, such as the one used in experiment 3, that depends entirely on placing chords within the global context of the tonality of the entire melody. If this were indeed the case, it would underscore the adaptive character of the reallocation of attention that appears to occur between 4 and 7.5 months of age. That is, the cognitive resources freed by inattention to irrelevant features in the environment can be devoted to more in-depth processing of relevant features. Though such a discussion is merely speculative as applied to this particular task, there are empirical examples from other domains that show an adaptive gain in one ability at the expense of another. Some of these, such as the loss of sensitivity to non-native phonetic contrasts in the service of efficient discrimination of native contrasts (Werker and Tees, 1984; Maye and Weiss, 2003), and the gain in ability to attend to relative pitch patterns at the expense of ability to attend to absolute pitch patterns (Saffran and Griepentrog, 2001; Saffran, 2003), have been discussed above. Critically, freeing processing resources is not the only benefit of inattention to less reliable cues. In the cases mentioned above, attention to absolute pitch information or to non-contrastive phonetic distinctions could actually obscure meaningful generalizations by causing a set of stimuli to appear more disparate than it would if irrelevant information was stripped away.

MUSICAL CORPUS SURVEY

The conclusion reached through the above experiments was that by 7.5 months of age have learned to ignore sequential dependencies of the AAB and ABA sort in their musical input, but attend to the identity of the last chord in each phrase with respect to its position in the key of the phrase. The claim was that this discrepancy is due to differing cue reliability in the natural music the infants are exposed to. That is, while a large proportion of phrases in natural children's music end in I or V, AAB and ABA sequences are not found with sufficient frequency to warrant attention to these kinds of relationships.

Up to this point, it has been a mere intuition that this discrepancy in cue reliability exists. In order to determine whether this intuition is correct, an empirical analysis of actual children's music was performed.

Methods

Source Materials

The songs used were drawn from an anthology of popular children's songs (Winn and Miller, 1966). From within this anthology, songs whose titles returned at least 100,000 Google hits were analyzed. Song titles for which none of the top five Google hits referred to the song itself were excluded due to the difficulty of accurately assessing their frequency. Most songs in this category had a single word, (e.g. "Raindrops"), or a single determiner-noun pair (e.g. "The Turkey"), for a title. This method yielded 22 songs.

Coding Method

The songs were broken into two-measure phrases. In several cases, a “pickup” to the next phrase occurred at the end of the second of the two measures. In these cases, punctuation or capitalization in the lyrics was used to determine where phrase boundary was located. Each phrase was coded for both the repetition pattern of the last three notes and for the chord symbol printed above the last note. Repetition patterns were AAA, AAB, ABA, ABB, and ABC; chord symbols were categorized as I, V, or other, based on the key signature of the song. All 22 songs were in a major key.

Results

The frequencies of each of the patterns of interest are shown in table 1. Of greatest interest, collapsing across all sequences with a repetition (everything but ABC), yielded 60 (43.5%) phrases across the four types. In contrast, 128 (92.8%) of the phrases ended in either I or V.

It is difficult to determine what a “chance” baseline frequency would be for each of these categories. An extremely naive approach would be to assume that each of the seven diatonic notes in the major scale (i.e., do, re, mi, fa, sol, la, ti) is equally likely at any position in the phrase, and that each of the seven diatonic chords (I, ii, iii, IV, V, vi, and vii^{o}) is equally likely to end a phrase. Using this naive method for generating phrases, it is straightforward to calculate that 1/7 of phrases would end in a I chord, and an additional 1/7 would end in a V chord. Therefore, 2/7, or 28.6% of phrases should end in I or V if they were generated according to this random scheme. The actual proportion of 92.8% is clearly much higher than this baseline.

A sequence of three notes with no repetitions (i.e., the ABC pattern) should be expected to occur 1 * 6/7 * 5/7 of the time, since the first note can be any note, the second note can be any of the 6 notes that the first note isn't, and the third note can be any of the 5 remaining notes. Thus, 61.2% of three-note sequences can be expected to have an ABC pattern, and hence 38.8% would have one of the other possible patterns. Note that this method of calculating a chance baseline produces quite a low baseline. The primary reason for this is that it does not take into account the "smoothness" of music, that is, that large leaps in pitch from note to note are quite unlikely. Rather, most melodies, especially simple folk melodies, proceed in a step-wise fashion, with a few small jumps of thirds and fourths. While the scheme used here treats notes an octave apart as identical, effectively constricting the range of a sequence to a single octave, in reality a three-note sequence is likely to have a smaller range even than that. However, even without taking this into account, the 43.5% of phrases that have a repetition is not much higher than the 38.8% baseline.

Chi-square goodness-of-fit tests were performed on data, using the baselines calculated above to calculate the expected frequencies. The proportion of phrases ending in I or V was revealed to be much higher than chance, $\chi^2(1) = 278.55$, $p<10^{-61}$. The distribution of repetition versus no repetition patterns was not significantly different from chance, $\chi^2(1) = 1.29$, $p>0.25$.

Discussion

It appears that, indeed, ending in I or V is a much more reliable cue in children's music than is ending in a pattern with a repetition, such as AAB or ABA. This result justifies the conclusion that the reason 7.5-month-olds learn a generalization involving the identity of the last chord in each phrase, but not one depending on positional sequential repetition, is that they have learned that sequential repetition is not a reliable cue in music, whereas ending in I or V is quite reliable.

GENERAL DISCUSSION

The three experiments discussed here, when taken together with the findings of Marcus, et al. (1999) and Marcus, et al. (2007), seem to reveal a cognitive reorganization, occurring in infants between 4 months and 8 months of age. The result is that the older infants no longer consider sequential dependencies among chords in a musical sequence to be particularly relevant for learning about musical structure, and hence do not easily learn a generalization employing such dependencies, even though they easily use the same dependencies to learn about strings of syllables. Experiment 3 showed that 7.5-month-olds can make generalizations about relational structure in musical sequences, provided the shared feature is one that is central to the structure of natural music. This result rules out the possibility that 7.5-month-olds' failure in the AAB/ABA chord task is due to a generalized loss in ability to make relational generalizations in music. These findings lend support to an intermediate view about the domain-generality of learning mechanisms. It need not be the case that learning mechanisms are either innately domain-specific or completely and eternally domain-general. Instead, it appears more likely that, at least in some cases, learning itself adapts to the structure of the environment.

This discussion is related, though not identical, to the long-standing debate in cognitive science about the modularity of cognition. Fodor (1984) famously advanced the hypothesis that much of cognition is modular in nature, where the defining features of modular processes were that they were informationally encapsulated (that is they communicated with other processes through very limited input and output channels), and that they were domain-specific. Since that proposal decades ago, the notion of modular cognition has become all but inseparable from the nativist school in cognitive science, which minimizes the role of experience in cognition, holding that genetic factors strongly specify the way in which cognitive processes function. The idea that much of cognition may be modular in nature does not logically imply a strong nativist view, however. Much as innate domain-specificity and eternal domain-generality are not the only two possibilities for learning mechanisms, cognitive modules might come to exist without having been built into the genome. Jacobs (1997; 1999), using computational models, explores a number of ways in which cognitive/neural modules might emerge as a result of brain activity, rather than through strict genetic developmental processes.

Computational Models

Several computational models exist that include mechanisms for attention shifting as a function of feature reliability, including some that consider relation-based features. Kruschke (2003, 2005) proposes several related models of attentional shifting in order to account for the phenomena of blocking and highlighting, which other learning models have difficulty with. Blocking refers to an apparently irrational behavior that arises when learners first learn an association between cue A and outcome X, then are exposed to instances of cues A and B together predicting outcome X. Not only do they not treat B as predictive of outcome X, but they subsequently have difficulty learning any association with B. Kruschke attributes inability to learn about B to learned inattention. In the related phenomenon of highlighting, learners first learn that the presence of cues A and B together predict outcome X. They then learn that cues A and D together predict outcome Y. Even though B predicts X and D predicts Y with perfect reliability, learners strongly prefer outcome Y when presented with B and D together. Furthermore, they prefer outcome X when presented with cue A alone. Highlighting, too, can be accounted for if learners are shifting attention away from A when it is presented with D, since they have already learned that A predicts X (and thus must be irrelevant or overwhelmed by D if A+D predicts Y). Kruschke's models, though they differ in particulars, are motivated by the connectionist principle that prediction error should be minimized. The general architecture of these models is such that input from cues is modulated by a layer of attentional units before reaching an output layer. Learning involves first updating the connection weights from cues to attention, and only then learning the particular mapping from cue to outcome.

Tomlinson and Love (2006) propose a model, BRIDGES, that is based on Kruschke's (1992) ALCOVE model, but which adds the ability to represent analogical relations in the input. Like ALCOVE, BRIDGES operates on the principle of predictive error reduction, shifting attentional weights to various input dimensions to emphasize those features or relations that best predict outcomes. In order to learn analogical relations, BRIDGES adopts the structure-mapping account of analogy (Gentner, 1983), which emphasizes parallel connectivity and one-to-one mapping of predicates and arguments. As such, two exemplars are perceived as similar if a one-to-one mapping of their elements can be constructed such that arguments playing the same role in a predicate can be mapped to each other. Among the empirical findings accounted for by

BRIDGES is that of Marcus, et al. (1999). BRIDGES represents a series of syllables with both positional and categorical information. That is, the sequence GA-GA-TI would be represented as containing three entities, GA_1, GA_2 and TI_1, where each entity possesses a feature denoting its serial position, and another feature denoting its abstract phonological membership (e.g. TypeOf(GA_1,GA). Upon encountering a second AAB string, say LI-LI-NA, BRIDGES would attempt to align each of the arguments in the representations of the two strings. Across all 16 strings presented in the empirical study, a stable one-to-one mapping can be constructed within the TypeOf predicate, and hence BRIDGES upweights the attentional parameter to this relation. As a result, novel strings such as KO-KO-GA are perceived as similar to the familiarization exemplars, whereas strings such as KO-GA-KO, lacking a stable one-to-one mapping that preserves parallel connectivity, is perceived as dissimilar.

Although the models discussed above do a good job of accounting for attentional modulation within a single learning episode, they cannot currently model stored attentional modes that can be accessed as a function of the domain of the input. A promising computational approach to capture this dissociation among domains might be Hierarchical Bayesian Models (HBMs) (Gelman, 2003). One nice feature of Bayesian approaches in general with respect to the question of domain-sensitive attentional modes is their ability to modulate distributions by conditioning on available data. Specifically, a Bayesian model of attentional shifting could condition the distribution of attention on other cues present in the input, resulting in different attentional distributions in the presence of different cues. *Hierarchical* models of Bayesian inference extends to multiple levels of abstraction the property of Bayesian models in general that cognitive models of the world should balance simplicity (i.e., a high prior probability) with good fit to the data (i.e., a high likelihood).

One recent HBM, developed by Kemp, Perfors and Tenenbaum (2007) addresses a learning problem with particular relevance to the present discussion. In a task by Jones and Smith (2002), children were presented with exemplars of two kinds of category: objects and substances. Across all tokens, size, shape and material varied. Within object categories, however, shape was constant, while material and size varied; within substance categories, material was constant while size and shape varied. Children exhibited domain-sensitive learning, generalizing from single exemplars of novel categories based on shape when the novel item was an object, and based on material when it was a substance. Kemp, et al. (2007) capture this same behavior using an HBM. At the highest level of abstraction, their model is equipped with knowledge that categories can fall into ontological kinds, but not how many kinds there are. A prior distribution over the number of kinds favors fewer divisions. The next level down contains parameters corresponding to feature variability within categories of a kind and feature variability across categories within a kind. The model contains parameters for each possible feature, here shape, size, material and solidity. Upon training the model with exemplars akin to those encountered by the children in Jones and Smith (2002), it correctly infers the presence of two kinds. For one kind, the parameter corresponding to shape variability within categories settles on a low value, while the parameters corresponding to material and size variability within categories are high. Similarly, the parameter for material variability within the second kind is low, while those for shape and size are high. For both kinds, the parameters for solidity variability are low for both the within- and between-category settings. Correspondingly, when presented with a novel exemplar that is solid, the first level of inference classifies it as the first (object) kind, and chooses a shape match as a category partner. When the novel exemplar is non-solid, the model chooses instead a material match.

The advantage of a hierarchical approach is that domains can be expressed in a straightforward way as higher-order categories – that is, "categories of categories" where domains contain information not only about the *values* of features (such as that the domain of "objects" tend to have the feature "solid"), but also about the *variability* of features (e.g. that the feature "shape" has low variability within categories in the domain of "objects"). When the learning mechanism in a hierarchical model is that of Bayesian inference, a great deal of information can be induced from data, including information not only about what categories exist and what their members are, but also about what domains there are, and what the properties of those domains are. Among the properties of domains that can be learned using these models is the variability of various features, and correspondingly, which features should be attended to in order to gain the most information about a novel set of input.

Open Questions and Future Directions

There remain several open questions directly pertaining to this research. First, while it appears clear that some sort of change is taking place between 4 and 8 months of age that leads to a loss of ease in learning the AAB generalization, the hypothesis that the change is attributable to the older infants' greater experience with music requires further testing. An alternative possibility is that a sort of "cognitive atrophy" takes place, such that active environmental support is required to maintain the ability to learn generalizations over certain features, and that the default outcome is that abilities are lost. If indeed the observed loss is due to infants learning about the structure of music, as opposed to a developmental pruning process, then 7.5-month-old infants should be expected to succeed at learning about sequential dependencies in a novel domain, in which they have not developed any expectations. Presumably, without a schema to guide attention, 7.5-month-olds will perform in the novel domain similarly to the way 4-month-olds perform in the musical domain. An experiment using non-periodic sound effects is currently underway to test this prediction.

It will also be important to know whether the decreased performance is due to discrete changes in the older infants' model of the musical world, or whether it is merely the case that attention to sequential dependencies is attenuated. Would the 7.5-month-olds be able to learn the generalization with additional familiarization, or is no amount of training sufficient at that stage in cognitive development? A related issue is what happens later in development. Will infants regain the ability to learn these generalizations at some point? If they do, what sorts of processes are responsible? As infants' models of the world become more sophisticated (and their representations more compact) will they begin to draw more parallels and analogical connections between domains?

The notion that experience continually reshapes not only knowledge and representations but the learning process itself invites revisitation of the concept of a critical period, particularly that which appears to exist in the domain of language learning. Purely biological accounts have long been advanced to explain the phenomenon that, past a certain age, it is much more difficult to master the grammar of a new language (or, in a few tragic cases, any language at all), and essentially impossible to develop a native-sounding accent. Some epigenetic process, for example, might shut off some sequence of genes at a certain point in development, or the developmental pruning of neural connections might render certain kinds of learning impossible after a certain age. However, another possibility is that previous experience with one’s native language interferes with learning of a new one. Certain features and relationships that are critical in deciphering the grammar of the new language may not be meaningful in one’s native language, and may be effectively exiled from consideration when it comes to learning new generalizations.

Summary and Conclusions

The discrepancy in the performance of 7.5-month-old infants on two different tasks involving generalizations about the relations present in musical input appears to reflect how reliably those relations appear in natural children’s songs. Successful performance by younger infants on the generalization involving the “unnatural” relation supports a learning-based account of the older infants’ difference in performance. When placed in the context of previous research by Marcus, et al. (1999), which shows that infants the same age as the older group succeed at learning the generalization that is "unnatural" in music when the same generalization applies to linguistic input, these results constitute evidence that infants are learning to allocate their attention differently when learning about input from different domains. This approach provides a promising middle ground between accounts of learning that posit a great deal of detailed, innate, domain-specific knowledge, on the one hand, and accounts that assume completely domain-general learning, on the other. The position taken here is instead that sensitivity to domains exists, but that higher-order learning processes can play a prominent role in shaping that sensitivity. This account would seem to lend itself to Hierarchical Bayesian Models, which allow for a series of increasingly abstract probability distributions over features of the world to be inferred from data at every level except the highest one. At that highest level, innate knowledge must come into play, though that knowledge can hopefully take a much simpler and more general form than that which is typically assumed by those who assume a rich base of knowledge that is completely independent of learning.

REFERENCES

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

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In JR Hayes, ed. *Cognition and the development of language.* New York: Wiley.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, Netherlands: Foris.

Cristia, & Seidl, (in press). Phonological features in infants' phonotactic learning: Evidence from artificial grammar learning. *Language, Learning and Development.*

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.

Fodor, J. (1984). *The Modularity of Mind.* Cambridge, MA: The MIT Press.

Gaskell, M., & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144-158.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis (2nd Edition)*. New York: Chapman & Hall.

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

Gerken, L. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, 98, B67-B74.

Gerken, L., & Bollt, A. (in press). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms. *Language Learning and Development.*

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249-268.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.

Gómez, R. L. & Gerken, L. A. (1999). 11-month-olds are sensitive to structure in an artificial grammar. *Cognition*, 70, 109-135.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, 16, 48-55.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins. *Cognition*, 78, B53-B64.

Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: A computational approach. *Psychonomic Bulletin and Review*, 4, 299-309.

Jacobs, R. A. (1999). Computational studies of the development of functionally specialized neural modules. *Trends in Cognitive Science*, 3, 31-38.

Johnson, E. K. & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, 5, 219-232.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111-116.

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10, 307-321.

Kruschke, J. K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review*, 99, 22-44.

Kruschke, John K. (2003). Attention in Learning. *Current Directions in Psychological Science*, 12, 171-175.

Kruschke, John K. (2005). Learning Involves Attention. G. Hougton (ed.) *Connectionist Models in Cognitive Psychology*. London: Psychology Press.

Lany, J., & Gómez, R. (2004). The role of prior learning in biasing generalization in artificial language learning. *Proceedings of the 26th Annual Conference of the Cognitive Science Society, Chicago, IL.*

Lany, J., Gómez, R., & Gerken, L. (2007). The role of prior experience in language acquisition. *Cognitive Science: A Multidisciplinary Journal*, 31, 481-507.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition,* 46, 53-85.

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., & Gerken, L. A. (2000). Learning phonemes without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development*.

Maye, J., & Weiss, D. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. *Proceedings of the 27th Annual Boston University Conference on Language Development*, 508-518.

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

Morgan, J., & Demuth, K., eds. (1996). *Signal to syntax*. Mahwah, New Jersey: Erlbaum.

Newport, E., & Aslin, R. (2004). Learning at a distance: Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-62.

Piston, W. (1969). *Harmony*. New York: Norton.

Rivera-Gaxiola, M., Silva-Pereyra, J. & Kuhl, P. K. (2005). Brain potentials to native- and non-native speech contrasts in seven- and eleven-month-old American infants. *Developmental Science*, 8, 162-172.

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language,* 47, 172-96.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science,* 12, 110-14.

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

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., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by adults and infants. *Cognition*, 70, 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: the structure of the input affects infants' use of absolute and relative pitch. *Developmental Science*, 8, 1–7.

Saffran, J., & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental Psychology,* 39, 484-94.

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology,* 39, 706-16.

Tomlinson, M., & Love, B. C. (2006). From pigeons to humans: Grounding relational learning in concrete examples. *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI-06)*, 199-204.

Toro, J. M., & Trobalon, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, 67, 867-875.

Trainor, L. & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 394-402.

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.

Winn, M., & Miller, A., eds. (1966). *The Fireside Book of Children's Songs*. New York: Simon and Schuster.