When to Hold and When to Fold: Detecting Structural Changes in Statistical Learning

Benjamin D. Zinszer (bdz107@psu.edu)
Center for Language Science
Departments of Psychology & Statistics, Penn State University
University Park, PA 16802 USA

Daniel J. Weiss (djw21@psu.edu)
Center for Language Science
Department of Psychology and Program in Linguistics, Penn State University
University Park, PA 16802 USA

Abstract

Studies of statistical learning have documented a remarkable sensitivity to structural regularities in both infants and adults. However, most studies of statistical learning have assumed a single underlying causal structure with uniform variance. In previous work in which two structures are presented successively, a primacy effect has been reported in which only the first structure is acquired. The present study explores the conditions under which such primacy effects are observed and learners are capable of acquiring both structures. We argue that learners can detect multiple structures by monitoring the consistency of the input.

Keywords: speech segmentation, statistical learning

Introduction

Over the past twenty years, research on language acquisition has been transformed by the finding that infant and adult learners can use rudimentary statistics to parse artificial speech streams (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). A large number of follow-up studies have replicated and extended the initial findings, determining that statistical learning is neither domain specific (e.g., Fiser & Aslin, 2002a; Kirkham et al. 2002), nor even restricted to humans (Hauser et al., 2001; Toro & Trobalón, 2005). The term statistical learning has consequently come to be associated with a wide range of phenomena that rely on implicit calculations based on distributional regularities in the environmental input.

The utility of these statistical learning experiments for simulating the early stages of language acquisition has been widely acknowledged. However, with few exceptions, the input to learners in statistical learning experiments has been characterized by a single, highly invariant statistical structure. This uniform-variance property of the input does not reflect the substantial variability inherent in natural language corpora due to shifts in topic, speaker, accent, and even language (in the case of bilingual acquisition). In some instances, variance in the input may signal to the learner that they are in a new context for which a different statistical structure must be learned (e.g., a language change), but in other cases this variation represents noise and should not trigger a new structural representation (e.g., hearing foreign-accented speech). Thus, the critical challenge confronting language learners is much like Piaget's description of the processes of assimilation and accommodation (Piaget, 1985). The learner must ultimately determine the number of causal models that best characterizes the input, resolving when a new causal model is required and when the existing model can account for the observed data.

There are at least two potential sources of information that may facilitate learners to detect that there has been a change in structure over time, which in turn may facilitate the formation of multiple representations (Gebhart, Aslin, & Newport, 2009). The first source of information is the availability of a contextual cue that is correlated with a particular statistical structure (e.g., Weiss, Gerfen, & Mitchel, 2009; Gebhart, Aslin, & Newport, 2009). The existence of such a cue could result in computations that are performed over a subset of the input and then compared across contexts. If the computations differ by some criterion, it would trigger the learner to form multiple representations to accommodate the inputs associated with each context. A second potential source of information for learners may be derived from monitoring the consistency of the input (Basseville & Nikiforov, 1993; see Gebhart, Aslin, & Newport, 2009). If the surface statistics are entirely consistent, the learner may conclude that the input likely has arisen from a single underlying structure. Conversely, if the variance in the surface statistics exceeds some criterion, then the learner may conclude that the underlying structure has undergone some change (see Gebhart, Aslin, & Newport 2009; Qian, Jaeger, & Aslin, 2012).

To date, only a few experiments have tested whether contextual cues facilitate the formation of multiple representations when multiple inputs are presented. In a study by Weiss, Gerfen, & Mitchel (2009), learners were presented with two artificial languages comprised of four words each, in which the words were defined solely by transitional probabilities. The languages were interleaved in two-minute intervals twelve times total. When the languages were presented in a single voice, only congruent language pairs were learned (ones whose statistics, when combined, yielded similar transitional probabilities to the languages presented in isolation). Incongruent languages (ones whose statistics were incompatible and yielded a higher noise level when combined) were only learned when a contextual cue was added such that one language was presented in one voice and the other in a second voice. Gebhart, Aslin, and Newport (2009) used a similar methodology, presenting learners with two five-and-a-half minute segments of incompatible languages presented consecutively (in the same voice). They reported a primacy effect in which the first language was learned at above chance levels, while the second language was not. However, learners succeeded in acquiring both languages if there was an explicit cue (informing the learners they would acquire two languages) in conjunction with a brief pause between streams. Also, tripling exposure time to the second language allowed learners to perform above chance in both languages, indicating that both languages could be acquired given sufficient exposure to the new language. Together, these results support the notion that the presence of a contextual cue differentiating the inputs can facilitate the formation of multiple representations, perhaps providing the learner with a more efficient route to successful acquisition.

To the best of our knowledge, no study to date has systematically investigated whether and how learners can form multiple representations by monitoring the consistency of the input alone. Arguably, some of the results from the aforementioned studies begin to address this issue, though the findings have not been easy to interpret (e.g., Weiss, Gerfen & Mitchel reported that repeated presentations of incongruent languages in the same voice resulted in no learning whereas the single presentations in Gebhart, Aslin, and Newport resulted in a primacy effect). In Experiment 1a, we set out to initially replicate the primacy effect of Gebhart and colleagues using their own languages. Subsequently, we manipulate both duration and language-switching parameters to determine the conditions under which learners can acquire both languages by monitoring the consistency of the input in the absence of contextual cues such as speaker voice or explicit instructions.

Experiment 1

In Experiment 1a, our goal was to replicate the primacy effect reported by Gebhart, Aslin, and Newport (2009) by presenting learners with two consecutive artificial languages with no explicit cue to the transition. We subsequently extend the study by manipulating the number of transitions between the two languages during the familiarization phase. Thus, in Experiment 1b, we control for the amount of exposure to each language while adding additional transition points (i.e., presenting four 2-minute 45-second blocks of each language versus two 5:30 blocks in Experiment 1a).

Methods

Participants Thirty-four undergraduate students were recruited from a Psychology 100 subject pool. Participants were divided into two conditions: 17 (11 female, 6 male) participants in Experiment 1a with a mean age of 19.6 years, and 17 (12 female, 5 male) participants in Experiment 1b with a mean age of 19.6 years. All participants were English monolinguals by self-report.

Languages The speech stream was composed of two languages, each consisting of sixteen trisyllabic words based on 12 unique CV syllables. These artificial languages were previously used in Gebhart, et al.'s (2009) segmentation experiment. Individually, the languages could be segmented by tracking the transitional probabilities (TPs) between syllables, with high TPs between syllables within a trigram (representing a word) and low TPs between syllables across different trigrams (representing word boundaries). See Figure 1 for an illustration of TP-defined words boundaries.

In both languages, two vowel frames and six consonants were used to define the trisyllabic words. The within-trigram transitional probability for syllables was 0.50. Words within the stream were randomly sequenced yielding a transitional probability of 0.25 between word-final and word-initial syllables. The second language rearranged the vowel frames and consonants of the first language, resulting in a syllable inventory that overlapped by 50%. The combined transitional probabilities (including all syllables across both languages) varied from 0.33 to 0.67 both within and between words. Consequently, they did not provide consistent cues for segmentation (see Figure 1).

Procedure Participants were instructed to listen to a brief recording of foreign speech and informed that they would later be quizzed on what they had heard. In Experiment 1a, participants listened to 5 minutes and 30 seconds of each language (produced in the same voice) consecutively without any cues to transition. Order of language

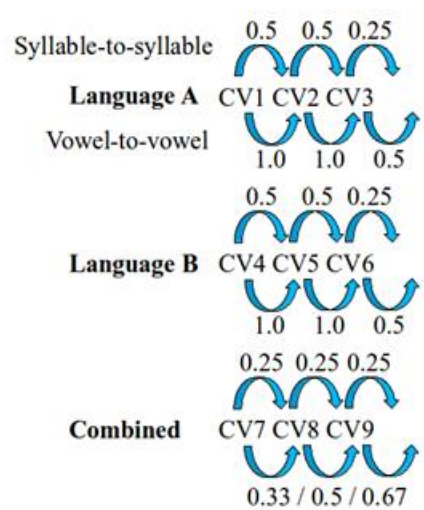

Figure 1 - Transitional probabilities defining the structure of each language. When combined, the TPs of each language result in a flat (uninformative) structure.

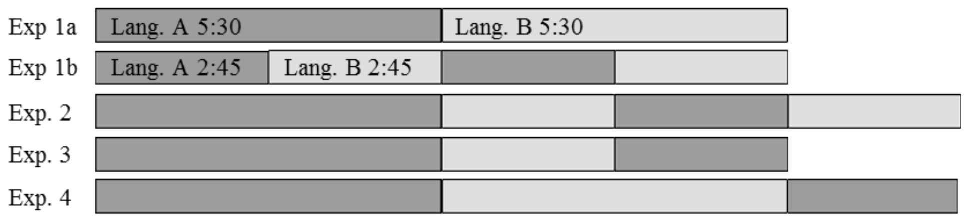

Figure 2 - Durations of each experiment depicted together. Dark bars represent Language A, and light bars represent Language B.

presentation was counter-balanced between participants, but for simplicity we will always refer to the Language A as the first language presented and Language B as the second. In Experiment 1b, participants listened to 4 consecutive blocks each consisting of 2 minutes and 45 seconds of one language (2:45 A + 2:45 B + 2:45 A + 2:45 B, order of actual languages was counter-balanced between participants). In both conditions, total exposure to each language was constant (5:30) as well as the total duration of the familiarization phase (11:00).

After familiarization, participants completed a test phase with thirty-two two-alternative forced choice trials in which participants selected between statistically-defined words and partwords. The partwords consisted of either the last syllable of a word followed by the first two syllables of another word or the last two syllables of a word followed by a single syllable of another word. These items occurred during the familiarization but were characterized as partwords since the within-trigram transitional probabilities were low. Participants were asked to judge which of the trigrams sounded more familiar, with statistically-defined words being counted as correct responses.

Results & Discussion

Mean correct responses on the test trials were computed for each language. In Experiment1a, participants scored a mean accuracy of 0.746 (*SD*=0.152) on Language A and 0.581 (*SD*=0.192) on Language B. These scores indicated a primacy effect in which accuracy on Language A significantly exceeded Language B (paired $t(16)=2.82$, $p=0.012$). Accuracy on Language A was significantly above chance ($t(16)=6.67$, $p<0.001$) while accuracy in Language B was not ($t(16)=1.73$, $p=0.102$). By contrast, in Experiment 1b, Language A and Language B did not significantly differ (paired $t(16)=1.00$, $p=0.331$). Also in contrast to Experiment 1a, Language A significantly exceeded chance (A: $M=0.673$, $SD=0.151$, $t(16)=4.73$, $p<0.001$), while Language B was also marginally significant ($M=0.603$, $SD=0.207$, $t(16)=2.05$, $p=0.057$). Results are illustrated in Figure 3.

In Experiment 1, we successfully replicated the primacy effect of Gebhart, et al. (2009) and discovered that increasing the number of switches between the languages could eliminate the primacy effect. Experiments 1a and 1b differed only in the duration of the individual exposure segments (5:30 vs. 2:45) and the number of switches between languages in the familiarization phase (1 vs. 3). Two causal hypotheses may be proposed for these results: First, the greater number of switches in the 1b stream may cue the listeners to the existence of two structures, allowing them to begin acquiring Language B. Alternately, in Experiment 1a (and the previous experiment by Gebhart and colleagues), learners may become entrenched in the statistical structure of Language A due to the lengthy duration of initial exposure. This entrenchment may inhibit detection or acquisition of the new structure. Experiment 2 was designed to disentangle these hypotheses.

Experiment 2

In Experiment 2, we contrasted the entrenchment and switching hypotheses proposed to explain the results of Experiment 1. We accomplished this by presenting participants with the following sequence of languages without any breaks in between: 5:30 of Language A followed by 2:45 of Language B, 2:45 of Language A again, and finally 2:45 of Language B. The entrenchment hypothesis predicts that the primacy effect found in Experiment 1a should also be present for Experiment 2 since the duration of the initial block of Language A is identical. The switching hypothesis predicts that both languages will be learned at significantly greater than chance levels since there are three transitions. This prediction of learning is somewhat counter-intuitive given that Language B was not learned in Experiment 1a and here we are increasing exposure to Language A.

Methods

Participants Twenty Psychology 100 students participated (12 female, 8 male; mean age 19.9 years).

Procedure As noted above, in this experiment the languages were configured as follows: 5:30 A + 2:45 B + 2:45 A + 2:45 B (see Figure 2). All other procedures were identical to Experiment 1.

Results & Discussion

No primacy effect was observed in Experiment 2, where Language A accuracy ($M=0.697$, $SD=0.161$) did not significantly differ from Language B accuracy ($M=0.694$, $SD=0.137$; paired $t(19)=0.07$, $p=0.943$). Moreover, both languages significantly exceeded chance performance

(A: $t(19)=5.46$, $p<0.001$; B: $t(19)=6.31$, $p<0.001$). Our results clearly reject the entrenchment hypothesis, lending support to the switching hypothesis as both languages were learned at above chance levels and performance did not significantly differ between languages.

Experiment 3

In Experiment 2, we eliminated the primacy effect found in Experiment 1a by increasing the number of switches between languages, even though Language A possessed a relative advantage in initial presentation duration and overall exposure time. In Experiment 3, we eliminate the last Language B exposure to test whether Language B may have been learned early in the sequence of exposures or whether the learning of B occurred only after the third transition (i.e., the last presentation). Notably, the only occurrence of Language B coincides with the first transition point in the sequence (and there are fewer transitions overall). Understanding when learning occurs may shed light on the type of processing that may be occurring for the unlearned language in conditions eliciting a primacy effect.

Methods

Participants Fifteen Psychology 100 students participated (13 female, 2 male; mean age 19.3 years).

Procedure In this experiment, languages were configured as follows: 5:30 A + 2:45 B + 2:45 A (see Figure 2 for illustration). All other procedures were identical to those described for Experiment 1.

Results & Discussion

The primacy effect emerged again in this experiment (paired $t(14)=2.73$, $p=0.016$), as Language A accuracy ($M=0.729$, $SD=0.179$, compared to chance: $t(14)=4.95$, $p<0.001$) significantly exceeded Language B ($M=0.546$, $SD=0.139$, compared to chance: $t(14)=1.28$, $p=0.221$). The contrast between these results and Experiment 2 highlights the importance of the second presentation of Language B for learning. In the absence of the third switch and additional exposure, performance was at chance levels for Language B. These results raise the question of whether the deficit in Language B learning was a function of the removal of the third language switch or the decrease in overall Language B exposure (from 5:30 to 2:45).

Experiment 4

The primacy effect observed in Experiment 3 emerged in the context of fewer switches and very short overall exposure. Therefore, in Experiment 4, we matched the overall exposure durations of Experiment 2 by providing learners with 5:30 of Language A followed by 5:30 of Language B and then an additional 2:45 of Language A again. This is essentially the sequence presented in Experiment 1a followed by an additional short block of Language A familiarization. Accordingly, the overall duration of each language presentation resembles Experiment 2 (in which both languages were learned), but here only two language switches are provided. Also like Experiments 2 and 3, Language A is advantaged in total exposure duration relative to Language B.

Methods

Participants Seventeen Psychology 100 students participated (13 female, 4 male; mean age 19.3 years).

Procedure In this experiment, languages were configured as follows: 5:30 A + 5:30 B + 2:45 A (see Figure 2 for illustration). All other procedures were identical to those described for Experiment 1.

Results & Discussion

Although Language A did not significantly exceed chance ($M=0.596$, $SD=0.210$, $t(16)=1.88$, $p=0.079$) and Language B significantly exceeded chance ($M=0.632$, $SD=0.132$, $t(16)=4.12$, $p<0.001$), accuracy in Languages A and B did

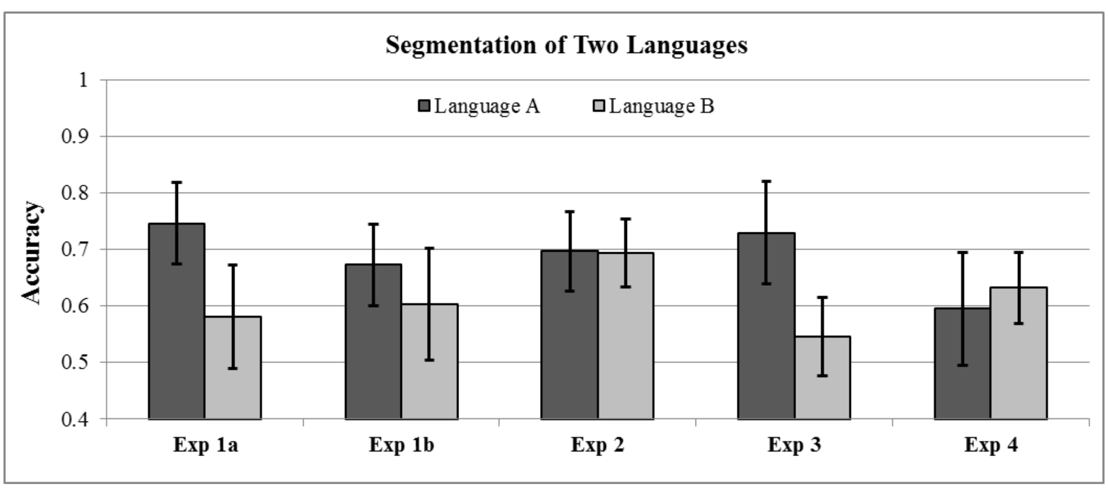

Figure 3 - Accuracy of participant responses in familiarity task. Chance level is 0.50, and error bars denote 95% confidence intervals of the mean.

not significantly differ from each other (paired $t(16)=-0.67$, $p=0.514$). See Figure 3 for illustration. The findings of Experiment 4 provide further evidence that that transitions between the languages alters learning. As in Experiment 2, we made the surprising observation that the additional exposure to Language A could facilitate learning of Language B. It is unclear at this point why Language A's learning was reduced and future experiments will explore the source of this effect. Irrespective of this pattern, the results from Experiment 4 do imply that some processing of Language B occurs even in conditions resulting in a primacy effect for Language A, such as the results reported by Gebhart, Aslin, & Newport (2009) and our Experiment 1a.

General Discussion

In the foregoing experiments, we have explored a range of conditions in which learners were familiarized with two artificial language streams characterized by incompatible underlying statistical structures. Unlike previous studies investigating statistical learning of multiple streams, no extralinguistic contextual cues were provided to the learners to signal the presence of a second language and facilitate learning. Consequently, successful learning of both languages relied on sensitivity to the structures themselves and the transition points between structures. In previous research, when statistically incompatible artificial languages were presented successively, learners have failed to successfully acquire both structures. In instances in which only a single switch was presented, learners exhibited a primacy effect (Gebhart, Aslin, & Newport, 2009) whereas when many switches occurred, there was a catastrophic interference effect in which no languages were learned (Weiss, Gerfen, & Mitchell, 2009; Mitchel & Weiss, 2010). We presented learners with the same languages used in the Gebhart, Aslin, & Newport (2009) study, and our results suggest that learners are sensitive to the transitions between the languages which can help them acquire both structures.

As noted above, in previous research, when learners receive input from two structures with only a single transition point, a primacy effect is observed (also replicated in Experiment 1a). In the original study, this effect could only be overcome by significantly extending exposure to Language B. Our findings have demonstrated that the primacy effect can also be overcome without increasing exposure at all. In Experiment 1b, exposure to Languages A and B were equivalent to the original study, though the languages were presented in smaller blocks and interleaved. This manipulation resulted in successful learning of both streams. Likewise, the results of Experiment 2 demonstrate that Language B can be learned even when our manipulation increases exposure only to Language A. That is, adding Language A training can, by virtue of the switching between languages, support learning in Language B as measured by our posthoc test.

We also explored whether there was any learning of Language B when it occurred after only a single transition. Gebhart, Aslin, and Newport (2009) as well as our Experiment 1a findings leave open the possibility that Language B was ignored altogether or that structures were unlearnable in light of the prior learning of Language A. The results of Experiment 4 cast doubt on either of these interpretations. In Experiment 4, learners acquired Language B at above chance levels despite the fact that Language B occurred in the exact same context as in the original condition (i.e., 5:30 of Language A followed by 5:30 of Language B). Like Experiment 2, the only manipulation in this experiment involved additional exposure for Language A.

We observed different learning outcomes between Experiments 3 and 4, suggesting that the amount of exposure to Language B prior to the second switch modulates the success of learning. It is possible that the 2:45 block of Language B in Experiment 3 did not provide adequate time for learners to sample the language, or perhaps 2:45 is insufficient to support learning (as we do not yet have baseline data for that duration). Because Experiment 3 is the only condition in the present study that limited Language B exposure to 2:45, further conditions will be necessary to explore this issue.

When three transition points are provided in the input (i.e., Experiment 1b and Experiment 2), the two structures become increasingly discriminable to learners, as evidenced by their above-chance performance in both languages. As noted above, the importance of these switches for detection of the second structure is highlighted in those experiments by the improvement in Language B performance despite only receiving additional exposure to Language A. We therefore conclude that learners are capable of identifying whether input streams contain one or multiple structures by monitoring the consistency of the input. This finding is in accordance with previous speculation regarding the conditions under which changes to statistical structures may be detected (Gebhart, Aslin, and Newport, 2009). The observed primacy effects in previous research and Experiment 1a cannot be attributed to entrenchment in the first language, as it has now been demonstrated that the primacy effects can be overcome with additional transition points between language streams.

While the present experiment made an extensive demonstration of language learning with only lingustic (syllable inventory) or statistical (TP) cues, a similar attempt has yielded markedly different results: Weiss, Gerfen, and Mitchell (2009) observed catastrophic interference when two incompatible languages were interleaved in 2 minute segments for a 24 minute stream. Under the switching hypothesis, we would have predicted significant learning of both languages. However Weiss and colleagues found that neither language was learned significantly better than chance. These two studies used different statistical structures, had different amount of overlap between the languages, and a different number of switches. Future experiments will try to systematically manipulate these parameters to better understand how overlap (and statistical compatibility) can influence the learning of multiple streams.

It has been hypothesized that describing how learners detect changes in statistical learning may be best explained by a hierarchical Bayesian model of change detection (Qian, Jaeger, & Aslin, 2012). How learners interpret non-uniform variance in statistical learning appears to rely on the availability of statistical and linguistic cues, such as changes in transitional probabilities or syllable inventory observed at a transition point between languages. These cues may lead learners to consider a second causal model to describe input (accommodation) over a single causal model under which the variance could occur (assimilation). This process of proposing causal models, weighting them by their likelihood, and comparing them to the input stream follows the procedure of Bayesian model comparison. In the case of extralinguistic cues in speech segmentation (e.g., pitch change or pause), this model comparison may be aided by the expectation of a context change and increase the prior probability of a two-model explanation of variance. Linguistic cues, such as the introduction of new syllables or the change in transitional probabilities between syllables may also effect such a change in the prior probability, though by themselves are insufficient. Our results suggest that change detection can be supported by variance-related events such as language switches, and provides further evidence that the Bayesian framework is a valuable analogy for statistical learning in multiple contexts.

Qian, Jaeger, and Aslin (2012) describe statistical cues to context change in terms of prediction error, i.e., a large deviation of the input stream from the learner's current model. Linguistic cues to speech segmentation may elicit such errors at language switches when the inventory or transitions between syllables change. This error-based cuing appears to be evidenced in the present study based on the importance of language switches to learners' performance. While previous research has demonstrated the utility of prolonged exposure to the second structure to detection of two contexts (Gebhart, Aslin, & Newport, 2009), we demonstrate that a relatively small set of high variation events can also increase the prior probability for a two-model hypothesis.

Our future work will attempt to determine the nature of processing that occurs during the unlearned streams. It is possible that learners detect the regularities in the second stream but discard it as noise, or that it is blocked by the learning of the first structure. Clearly, some information is gathered during those periods, as evidenced by the results of Experiments 2 and 4. One set of studies underway introduces a third structure into the sequence (either a new learnable artificial language or an unlearnable non-adjacent transitional probability language). Thus, the sequence is 5:30 of A followed by 2:45 of Language C (noise or learnable) followed by 2:45 of A and 2:45 of B. This condition tests whether switches by themselves are useful (without supporting the statistics of Language B). We are also currently engaged in neuroimaging studies to localize and contrast the learning of Languages A and B in a variety of conditions.

Acknowledgements

We wish to thank Dr. Richard Aslin and Timothy Poepsel for comments and discussion, as well as Tzuyu Chang for assistance with data collection. This project was funded by NIH grant R01 HD067250-01 to DJW.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science, 9*(4), 321-324.

Basseville, M., & Nikiforov, I. (1993). *Detection of abrupt changes – Theory and application*. Englewood Cliffs, NJ: Prentice-Hall.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822-26.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*, 1087-1116.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), 53-64.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), 35-42.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development (Vol. 985)*. Chicago: University of Chicago Press.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology, 3*(July), 228.

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

Toro, J. M., & Trobalon, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics, 67*(5), 867-875.

Weiss, D. J., Gerfen, C., & Mitchel, A. (2009). Speech Segmentation in a Simulated Bilingual Environment: A Challenge for Statistical Learning? *Language Learning and Development*, *5*(1), 30-49.