Title
Processes Underlying Infants' and Adults' Visual Statistical Learning

Permalink
https://escholarship.org/uc/item/8b44f93n

Author
Slone, Lauren

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Processes Underlying Infants' and Adults' Visual Statistical Learning

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Lauren Elizabeth Krogh Slone

2015

ABSTRACT OF THE DISSERTATION

Processes Underlying Infants' and Adults' Visual Statistical Learning

by

Lauren Elizabeth Krogh Slone

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2015

Professor Scott Pratt Johnson, Chair

How do learners make sense of their intricately structured visual and auditory environments? One important learning mechanism available to both infants and adults is the ability to detect statistical structure within visual and auditory inputs. Despite the scope of this "statistical learning" (SL) ability and the large literature that now surrounds it, the processes underlying SL, and how those processes may differ across modalities and change across development, remain unclear.

The goal of this dissertation was to clarify these mechanistic questions. Paper 1 reviews previous work on visual and auditory SL across development, highlighting the debate concerning the domain-generality of SL. Paper 2 investigates the domain-generality of infants' SL by employing a visual SL task comparable to previous auditory SL tasks. Findings from four experiments with 8-month-olds suggest that visual SL is constrained compared to auditory SL in infancy. Individual differences in visual SL performances were unrelated to visual short-term

memory performances, but were related to overall cognitive ability. These findings suggest that visual SL ability is important for, and constrained by, early cognitive development.

Papers 3 and 4 examine the representations that result from visual SL. I evaluate the ability of two major classes of SL models – statistical models and chunking models - to correctly predict adults' and infants' performances on visual SL tasks. Across five experiments, adults and 8-month-old infants discriminated between high- and low-probability visual sequences, providing strong evidence of SL. Critically, adults represented pairs of items embedded within larger sequences, but did not represent so-called "illusory sequences." These results support the competitive chunking model of Servan-Schreiber and Anderson (1990), and suggest that adults represented visual sequences in terms of hierarchical levels of chunks. In contrast, 8-month-olds did not show evidence of representing embedded or illusory sequences, suggesting that infants represented visual sequences in terms of only the highest level of chunks. Together, these studies not only suggest that the representations that result from visual SL are best captured by chunking models, but also that the type of chunking learners engage in may change between infancy and adulthood.

The dissertation of Lauren Elizabeth Krogh Slone is approved.

Gregory A. Bryant

Alan Dan Castel

Hongjing Lu

Catherine M. Sandhofer

Scott Pratt Johnson, Committee Chair

University of California, Los Angeles

2015

For my God,

my savior, comforter, and counselor.

Thank you for giving me new life and a reason to live it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank all of the colleagues, friends, and family who have supported me throughout graduate school and helped to make this time such a fun and rewarding experience. At a place as supportive, collaborative, and genuinely caring as UCLA, there are far too many people to thank. But first and foremost, I thank my advisers, Scott and Cathy. Scott, you have played a crucial role in this achievement. You have helped me to think critically and ask important questions, and you have been a model of determination and success. Thank you for always being available to me, for supporting me, and for helping me to become a confident and knowledgeable young researcher. Cathy, thank you for your continued support and mentorship. Working with you has opened up new populations and new lines of research to me, as well as opportunities for collaboration and publication that have benefitted me tremendously. Your expertise and mentorship have been incredibly influential in my research success and professional development, and have provided an excellent model for myself as a mentor.

In addition to my primary graduate advisors, I would like to thank two other faculty members who have been central to my success: my undergraduate advisor, David Rakison, and my collaborator, David Moore. David Rakison, thank you for getting me started on this career path. If it were not for your mentorship, I would not be where I am today. I am forever grateful to you. David Moore, thank you for being a supportive mentor and enthusiastic research collaborator throughout graduate school. I have benefitted incredibly from your expertise and encouragement.

I would also like to thank the faculty and graduate students of the Developmental area, as well as my professors from the Cognitive area, and my dissertation committee members. The collaborative and collegial nature of this department has no doubt played an important role in my

success. Thank you all for challenging me to aim high, for helping me to think more broadly, and for inspiring me to be more daring and confident in my research. Thank you especially to Heejung Park and Lauren Sherman for being wonderful colleagues as well as fabulous friends and roommates. I am also grateful to the past and present members of the Baby Lab and Language and Cognitive Development Lab, especially Haley Vlach, Elizabeth Goldenberg, Natsuki Atagi, and Bryan Nguyen. Haley, Elizabeth, and Natsuki, thank you for providing valuable feedback and encouragement over several years and throughout every step of the research process; I cannot imagine navigating academia without you. Bryan, thank you for patiently and precisely programming so many studies for me, allowing me to be so productive.

Thank you also to all of the research assistants who have generously volunteered their time and energy to make this research possible, including Maggie Anderson, Pearl Aquino, Elena Arciniega-Hernandez, Vera Chan, Melissa Chaung, Anna Chen, Laura Colosimo, Arielle Creager, Esmeralda Garrido, Kristen Gonzalez, Asya Grigoryan, Lana Halavi, Cynthia Huang, Jennifer Jarson, Brianna Kaplan, Eunyoung Lee, Leanna Maaz, Diane Motamed, Nour Nabi, Lyndsie Okumura, Tiffany Ou, Maribel Ramos, Ana Sanchez, Kortney Sasaki, Tatiana Serrano, Justina Taban, Hasmik Tokadjian, Karina Vich, and Brittany Zambrano.

Importantly, I would also like to thank the incredibly supportive family and friends who have encouraged me throughout graduate school. Thank you for giving me strength in times of stress and so much joy in times of success. To my husband Jeff, thank you for your love, your sense of humor, and for helping keep me sane. Dad, thank you for always encouraging my excitement for learning. Mom, thank you for your unending support and for being my model of selflessness, perseverance, and strength of character. Kris, you continue to be my rock and safe haven in times of stress – thank you for truly loving me unconditionally. Grandma, I miss you,

but I am so grateful for all that you taught me. Most importantly, I praise and thank my savior and Lord, Jesus Christ, for His grace in all things. Without Him, all of this would be meaningless.

VITA

EDUCATION

2011   M.A. in Psychology
University of California, Los Angeles

2009   B.S. in Biological Sciences and Psychology
Carnegie Mellon University
University and College Honors, Dean's List High Honors, Overall GPA: 4.0/4.0

SELECTED GRANTS AND FELLOWSHIPS

2012-2015   Graduate Research Fellowship, National Science Foundation

2011   Graduate Summer Research Fellowship, UCLA

2010-2012   Graduate Research Fellowship, Jacob K. Javits Foundation

2010   Graduate Research Mentorship Fellowship, UCLA

2009   Distinguished University Fellowship, UCLA

2009   Chancellor's Award, UCLA

2009   Senior Honors Thesis Grant, Carnegie Mellon University

2008, 2009   Small Undergraduate Research Grant, Carnegie Mellon University

ACADEMIC AND PROFESSIONAL EMPLOYMENT

2012-2015   Writing Consultant
Graduate Writing Center, University of California, Los Angeles

2013   Teaching Assistant, Psychology 120B, Sensation and Perception
University of California, Los Angeles

2011   Teaching Assistant, Psychology 134A-B, Applied Developmental Psychology I-II
University of California, Los Angeles

2011   Teaching Assistant, Psychology 101, General Psychology Lab
University of California, Los Angeles

2010   Teaching Assistant, Psychology 134F, Infant Care and Development
University of California, Los Angeles

SELECTED PUBLICATIONS

Slone, L. K. & Johnson, S. P. (2015). Statistical and chunking processes in adults' visual sequence learning. In R. P. Cooper (Ed.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Pasadena, CA: Cognitive Science Society.

Slone, L. K. & Johnson, S. P. (2015). Infants' statistical learning: 2- and 5-month-olds' segmentation of continuous visual sequences. *Journal of Experimental Child Psychology, 133,* 47-56. DOI:10.1016/j.jecp.2015.01.007 NIHMSID: NIHMS670651

Krogh, L., Vlach, H., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Language Sciences, 3,* 598. DOI:10.3389/fpsyg.2012.00598

Rakison, D. H. & **Krogh, L.** (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science, 15*, 43-53. DOI:10.1111/j.1467-7687.2011.01096.x

SELECTED PRESENTATIONS

Slone, L. K. & Johnson, S. P. (2015). *Statistical and chunking processes in infants' and adults' visual statistical learning*. Paper to be presented at the International Conference on Interdisciplinary Advances in Statistical Learning, San Sebastian, Spain.

Slone, L. K., Kim, H. I. & Johnson, S. P. (2015). *Infants track transitional probabilities in continuous sequences of faces but not shapes*. Poster presented at the Society for Research in Child Development Biennial Conference, Philadelphia, PA.

Slone, L. K. & Johnson, S. P. (2013). *Constraints on statistical learning in infancy*. Paper presented at the Eighth Biennial Meeting of the Cognitive Development Society, Memphis.

Krogh, L., Moore, D. S., & Johnson, S. P. (2013). *Infants' visual-manual object exploration and mental rotation performance.* Paper presented at the Society for Research in Child Development Biennial Conference, Seattle, WA.

Krogh, L. & Sandhofer, C. M. (2013). *Consider the category: The effect of spacing depends on the material to be learned.* Poster presented at the Society for Research in Child Development Biennial Conference, Seattle, WA.

Krogh, L. & Rakison, D. H. (2009). *The effect of action on causal perception in 3- and 4 1/2-month-old infants.* Paper presented at the Annual Meeting of the Minds Undergraduate Research Symposium, Pittsburgh, PA. Department of Psychology Talk Prize.

SIGNIFICANCE OF THE PRESENT RESEARCH

A central question of cognitive science and developmental psychology concerns how learners make sense of their intricately structured visual and auditory environments. Of particular interest to developmental psychologists is how infants acquire so much information so quickly and seemingly effortlessly in the first years after birth. One important learning mechanism available to infants as well as adults is the ability to detect statistical relations among elements in a sensory array. This "statistical learning" ability was first demonstrated with linguistic information (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b), but has since been found to operate across a variety of domains and modalities (e.g., Fiser & Aslin, 2001, 2002; Saffran, Johnson, Aslin, & Newport, 1999). Yet despite the scope of statistical learning and the large literature that now surrounds it, the processes underlying statistical learning remain unclear (Frank, Goldwater, Griffiths, & Tenenbaum, 2010). Moreover, it is unclear whether all forms of statistical learning reported across domains and across development rely on the same underlying processes, as might be implied by the fact that all are referred to by the same term "statistical learning" (Thiessen, Kronstein, & Hufnagle, 2013).

The goal of this dissertation is to clarify this mechanistic question. My strategy is to employ a single method (a sequential visual statistical learning task) to extend previous research on visual statistical learning in both infants and adults and to investigate the fit of statistical learning models to human performance. The organization of this dissertation is as follows. I first review previous work on auditory and visual statistical learning across development, highlighting the debate concerning the domain-generality of statistical learning and the need for statistical learning research that employs comparable tasks across domains (Paper 1). Next, I investigate the domain-generality of infants' statistical learning by employing a visual statistical learning

1

task comparable to that used to examine auditory statistical learning (Paper 2). Specifically, I examine: (1) whether infants can acquire via visual statistical learning the same complexity of statistical structures that infants acquire via auditory statistical learning, and (2) constraints on infants' visual statistical learning. Finally, I investigate the representations learners store during visual statistical learning by evaluating the ability of two major classes of statistical learning models – statistical models and chunking models - to correctly predict adults' (Paper 3) and infants' (Paper 4) performances on visual statistical learning tasks.

My approach is innovative in its use of (1) a single method of investigating statistical learning that is suitable for participants ranging from less than 2 months of age to adults, and (2) multiple measures of learning. Moreover, there are six contributions of this work: (1) it reviews statistical learning research across domains and across development, (2) it introduces a variety of new human statistical learning data, (3) it suggests important differences in infants' learning of visual, compared to auditory, statistical structures, (4) it identifies potential constraints on infants' visual statistical learning ability, (5) it suggests that both infants' and adults' visual statistical learning performances are best captured by chunking models, and (6) it identifies differences in the types of chunking infant and adult learners engage in.

References

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

Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning Memory and Cognition* 28, 458-467.

Frank, M.C., Goldwater, S., Griffiths, T., and Tenenbaum, J.B. (2010). Modeling human performance in statistical word segmentation. *Cognition,* 117, 107-125.

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

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

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

Thiessen, E.D., Kronstein, A.T., and Hufnagle, D.G. (2012). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin.*

PAPER 1:

Statistical learning across development: Flexible yet constrained

Abstract

Much research in the past two decades has documented infants' and adults' ability to extract statistical regularities from auditory input. Importantly, recent research has extended these findings to the visual domain, demonstrating learners' sensitivity to statistical patterns within visual arrays and sequences of shapes. In this review we discuss both auditory and visual statistical learning to elucidate both the generality of and constraints on statistical learning. The review first outlines the major findings of the statistical learning literature with infants, followed by discussion of statistical learning across domains, modalities, and development. The second part of this review considers constraints on statistical learning. The discussion focuses on two categories of constraint: constraints on the types of input over which statistical learning operates and constraints based on the state of the learner. The review concludes with a discussion of possible mechanisms underlying statistical learning.

Statistical Learning Across Development: Flexible yet Constrained

To survive, an organism must be capable of organizing and interpreting the constant stream of sensory input it receives. Research in the last two decades has revealed powerful statistical learning abilities in infants and adults, including the developing capacity to extract statistical regularities from a variety of auditory inputs including artificial and natural language (e.g. Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Pelucchi, Hay, & Saffran, 2009) and non-linguistic auditory stimuli (Saffran, Johnson, Aslin, & Newport, 1999). An independent line of research has extended these findings to the visual domain, demonstrating infants' and adults' sensitivity to statistical patterns within visual arrays and sequences of shapes (e.g. Bulf, Johnson, & Valenza, 2011; Fiser & Aslin, 2001, 2002a, 2002b; Kirkham, Slemmer, Richardson, & Johnson, 2007; Kirkham, Slemmer, & Johnson, 2002).

The current review discusses auditory and visual statistical learning to elucidate both its generality and its constraints. We first outline the major findings of the statistical learning literature with infants, followed by discussion of statistical learning across domains, modalities, and development. The second part of this review considers constraints on statistical learning. The discussion focuses on two categories of constraint: constraints on the types of input over which statistical learning operates, and constraints based on the state of the learner. The review concludes with a discussion of possible mechanisms underlying statistical learning.

Auditory Statistical Learning

Artificial Language

Given the richness and complexity of a natural language, how is it that infants acquire vocabulary and structure so rapidly, and seemingly effortlessly, in their first years after birth? For example, one challenge facing young language-learners is the fact that speakers do not mark

word boundaries with pauses, and listeners must rely on other information to accomplish this task. Early in the "cognitive revolution," researchers hypothesized that the statistical structure of language might be important for word segmentation (Harris, 1955; Hayes & Clark, 1970). For instance, Hayes and Clark (1970) tested adults' ability to segment "words" from a continuous stream of speech analogs in which the only cue to word boundaries was the distribution of the phonemes. Adult participants successfully segmented words, suggesting sensitivity to statistical information in speech. However, Hayes and Clark did not specify a mechanism to account for this result.

Building upon these findings, Saffran and colleagues (Saffran et al., 1996a, 1996b) proposed a mechanism for statistical word segmentation: transitional probability (TP) detection. In their experiments, adults, first-graders, and 8-month-olds were presented with a continuous stream of speech from an artificial language in which word boundaries were indicated by differing TPs between syllables within words (high TPs) and across word boundaries (low TPs). After brief exposure to this language, listeners in all three age groups were able to distinguish between high TP syllable sequences ("words") and low TP sequences ("part-words"). Thus, both infant and adults learners appeared sensitive to the TP information contained in the speech stream, suggesting that statistical learning via sensitivity to TPs is a possible mechanism contributing to language acquisition.

Although such early studies in infant statistical learning conceptualized statistical learning as sensitivity to a particular conditional relation, transitional probability, more recent research highlights a variety of other conditional statistics (e.g., mutual information) that could be used to distinguish words from foil items. This point is discussed in greater detail in a subsequent section, however we mention it briefly here to point out that, although several studies

are described in terms of differing TPs, it remains unclear which conditional relations participants rely upon to segment sequences.

One limitation to the design of the aforementioned studies was that frequency information covaried with conditional probability statistics. That is, high TP words occurred more frequently than low TP part-words in the learning (familiarization) phase of the experiment, and it remained unclear whether participants distinguished syllable sequences based on differences in conditional relations or simply differential frequencies of occurrence during learning. To address this issue, Aslin, Saffran, and Newport (1998) conducted a "frequency-balanced" version of their original study, with words and part-words appearing equally frequently, such that only sensitivity to conditional relations could be used to distinguish the two types of sequences. Aslin et al. found that 8-month-old infants were still able to distinguish high and low TP sequences. This result suggests that infants can track conditional probability information independent of co-occurrence frequency and use this information to determine word boundaries. Taken together, this work demonstrated the potential for statistical learning to support early language acquisition.

The possibility that statistical learning is a primary mechanism underlying early language acquisition raises the question of the age at which statistical learning is functional in young infants. Teinonen, Fellman, Näätänen, Alku, and Huotilainen (2009) examined statistical learning in sleeping newborns by presenting a continuous stream of 3-syllable words in an artificial language similar to that employed by Saffran et al. (1996a), in which the only cues to word boundaries were the conditional relations or frequencies of co-occurrence between syllables. Using electroencephalography, they measured newborns' event-related potential (ERP) negativities to the first, second, and third syllables in the words. Teinonen et al. (2009) found a

significant difference between the ERP negativity to the first and third syllables, indicating that the neonatal brain is sensitive to word boundaries marked by conditional relations and reacts differently during word onset compared to word offset. This research demonstrates, therefore, that statistical learning is functional even in newborn infants, and perhaps contributes to language acquisition even prior to birth.

For statistical learning to be a primary mechanism underpinning infants' early language acquisition, however, it must be able to scale up to the demands of more complex natural language (Johnson & Tyler, 2010). The aforementioned studies employed artificial speech composed entirely of bisyllabic words or entirely of trisyllabic words. Natural language, in contrast, consists of much more varied word types. To simulate more natural language learning, Johnson and Tyler (2010) investigated infants' ability to segment an artificial language composed of both bi- and trisyllabic words. Interestingly, neither 5.5- nor 8-month-old infants were able to segment this language, suggesting that certain characteristics of natural language, such as varied word length, may make segmentation more difficult compared to segmentation of artificial languages.

Other research, however, suggests that some characteristics of natural language may help to make statistical word segmentation possible. For instance, Thiessen, Hill and Saffran (2005) found that 7-month-olds were able to segment an artificial language containing words of varying length when the language was produced with infant- but not adult-directed prosody. As an artificial language becomes more complex (here, by consisting of words of mixed, as opposed to uniform, length), therefore, other natural speech cues such as exaggerated prosody may be needed to facilitate statistical word segmentation.

Indeed, conditional probabilities have never been posited as the sole cue to word

segmentation in natural language. Instead, researchers have suggested that initial sensitivity to conditional probabilities may facilitate language acquisition by bootstrapping sensitivity to other linguistic cues. For instance, in English, lexical stress serves as a cue to word boundaries as a majority of English words are stressed on their first syllable (Thiessen & Saffran, 2003). Statistical segmentation mechanisms may facilitate sensitivity to stress cues by providing infants with an inventory of words from which they can discover the dominant stress pattern of their native language (Swingley, 2005; Thiessen & Saffran, 2003, 2007).

In the next section, we discuss research that provides even stronger support for the possibility that statistical learning contributes to language acquisition by examining infants' statistical learning in natural language.

Natural Language

The aforementioned research focused on statistical learning in the context of synthesized artificial languages. More recent research has examined more natural language learning contexts, such as sequences of grammatically correct and semantically meaningful sentences in natural speech. Pelucchi, Hay, and Saffran (2009) examined 8-month-olds' ability to extract statistical regularities from an unfamiliar natural language (Italian for English-learning infants). Infants were presented with a constant stream of fluent infant-directed Italian speech for approximately two minutes. After this brief exposure, infants provided evidence of discrimination between high- and low-TP bisyllabic words. Importantly, both types of words had occurred equally frequently in the speech stream, indicating that infants were using conditional probability information, not simply frequency information, in discriminating between words.

The Pelucchi et al. (2009) results imply that infants discriminated likely from unlikely sound sequences in natural language, but they leave open the critical question of how learners

represent extracted statistical information. Saffran (2001) took an important step in addressing this question by asking whether English-learning infants treat segmented syllable sequences as candidate English words or simply as highly probable sound sequences. In this experiment, 8-month-old infants were familiarized to a continuous stream of artificial speech composed of nonsense words similar to those used in Saffran et al. (1996a). Following familiarization to the stimuli, infants participated in a post-familiarization test. This test compared infants' listening time to speech in which words and part-words were embedded in either simple English (e.g. "I like my *tubido*") or matched nonsense (e.g. "zy fike ny *tubido*") frames. If infants treated the outputs of statistical learning simply as highly probable sound sequences, both the English and nonsense frame conditions should have elicited similar listening preferences. However, if infants treated the outputs of statistical learning as candidate English words, then they should have shown differential listening preferences when those units were embedded in English versus nonsense frames. Saffran found that infants exposed to English frames listened significantly longer to words in this English context than to part-words, and that this difference in listening preference for words versus part-words did not extend to the nonsense frame condition. These results suggest that the statistical learning mechanisms underlying word segmentation do generate word-like units and raises the question of whether these units are available to support other aspects of language acquisition, such as mapping words to meaning.

Establishing a link between sound and meaning is an essential aspect of language acquisition, particularly for young language learners. Graf Estes, Evans, Alibali, and Saffran (2007) investigated the connection between statistical word segmentation and object-label learning in 17-month-olds. Infants were presented with 2.5 minutes of fluent speech composed of bisyllabic nonsense words where the only cues to word boundaries were the conditional relations

11

between syllables. Immediately following this segmentation task, infants were habituated to two object-label combinations, presented one at a time. For each combination, infants heard a bisyllabic sound sequence from the segmentation task while viewing a 3D object on a computer screen. For half the infants, the bisyllabic sound sequences were words from the segmentation task, and for the other half, the sound sequences were non-words (Experiment 1) or part-words (Experiment 2). Following habituation to these two object-label pairings, infants were presented with two types of test trials. "Same" test trials presented the same object-label combinations from the habituation phase. "Switch" test trials switched the labels for the two objects such that the label for object 1 was played while the infant viewed object 2. Longer looking on switch trials would suggest that infants were sensitive to the change in word-object pairings and was therefore taken as evidence of acquisition of the object-label associations. Graf Estes et al. found that only infants exposed to words from the segmentation task as object labels looked longer on switch compared to same test trials. This indicates that by 17 months of age, infants may be able to map newly segmented sound sequences ("words") to novel objects as linguistic labels, but are unable to do so with non-words or part-words. These results support the claim that statistically segmented sound sequences are word-like and suggest that the output of auditory statistical learning is represented linguistically.

Recent work has also found associations between statistical learning abilities and natural language processing (Conway, Bauerschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2012). For instance, Misyak and Christiansen (2012) found that even after controlling for measures of short-term and working memory, vocabulary, reading experience, cognitive motivation, and fluid intelligence, performance on statistical learning tasks was the key predictor of comprehension of natural language sentences. Such findings suggest that statistical learning

may be relevant to language learning not only because extracted statistical information may be represented linguistically, but also because statistical and language learning might overlap in their underlying mechanisms (Christiansen, Conway, & Onnis, 2007; Misyak & Christiansen, 2012; see also work on cross-situational statistical learning, e.g., Smith & Yu, 2008).

Non-linguistic Stimuli

Demonstrations that conditional probability information extracted from auditory input is represented linguistically (Graf Estes et al., 2007; Saffran, 2001) and that learners form associations between auditory statistical learning and language learning (Conway et al., 2010**;** Misyak & Christiansen, 2012) raise the question whether statistical learning is language-specific, or whether it also operates over non-linguistic stimuli. In the auditory domain, Saffran, Johnson, Aslin, and Newport (1999) found that both infants and adults appeared to detect statistical regularities in nonlinguistic sequences of "tone-words." The procedure and stimuli used were modeled directly after those used in Saffran et al.'s (1996a, 1996b) studies employing speech, allowing for a direct comparison of participants' performance with tones and syllables. Both adults and infants performed with similar accuracy in discriminating words from part-words, regardless of whether these units were instantiated in syllables or tones. These findings suggest that statistical structure can be extracted from auditory input regardless of the domain in which it is presented (syllables or tones), and raise the possibility that statistical learning might also function over input from other modalities.

<center>**Visual Statistical Learning**</center>

Investigating infants' and adults' extraction of statistical structure in visual input addresses the question of domain generality by asking whether or not statistical learning is limited to auditory input.

Infants

Kirkham, Slemmer, and Johnson (2002) examined infants' detection of statistical regularities from sequentially presented visual information. Two-, 5- and 8-month-old infants were habituated to a continuous stream of six looming colored shapes presented one at a time with no breaks or pauses between shapes. The six shapes were organized into three pairs that were presented in random order such that the boundaries between pairs were defined by TPs (TP=1.0 within pairs, TP=0.33 between pairs). Following habituation, infants viewed six test displays alternating between the familiar habituation sequence and a novel sequence composed of the same six shapes from habituation presented in random order. Infants at all three ages exhibited a significant novelty preference, suggesting that the infants were sensitive to statistical regularities that defined the visual shape sequences. This was the first published experiment to demonstrate not only infants' sensitivity to statistically defined structure in visual sequences, but also to suggest that statistical learning is a domain general learning process, capable of identifying statistical structure across modalities.

The Kirkham et al. (2002) study was also the first to investigate the developmental time-course of visual statistical learning during the first year after birth. Kirkham et al. found no significant differences in novelty preferences between age groups. This lack of observed development, combined with the finding that statistical structures could be detected after only a few minutes of exposure, suggests visual statistical learning may be functional at or soon after the onset of visual experience. Bulf, Johnson, and Valenza (2011) explored this possibility by investigating whether infants are capable of extracting statistical regularities from visual sequences at birth. Bulf et al. employed a habituation design similar to that used by Kirkham et al. (2002), presenting newborn infants (mean age 38 hours) with continuous sequences of either

four or six looming shapes following a statistically defined structure. Newborns provided evidence of detecting the structure of the shape sequences, though only in sequences composed of four, not six, shapes. Thus, statistical learning appears to be functional at birth, operating over both auditory (Teinonen et al., 2009) and visual input (Bulf et al., 2011), but is constrained, an issue we discuss in greater detail in a subsequent section.

The method of testing employed by Kirkham et al. (2002) and Bulf et al. (2011) demonstrated that infants can discriminate between structured and random sequences. However, it did not indicate what statistical or structural features allowed infants to make this discrimination. Rather than computing conditional statistics, as has been found in studies of auditory statistical learning, infants could have been responding to a variety of other features, such as frequency of shape co-occurrence, which co-varied with conditional probability information. Determining which features infants are sensitive to is important for understanding the extent and utility of statistical learning as detection of different statistical features allow varying degrees of associative learning and inference. For instance, *co-occurrence statistics* inform the observer of the likelihood of two events occurring together, but leave the observer uncertain of the likelihood of an event occurring given that the other has taken place. In contrast, *conditional probability statistics* serve to reduce uncertainty by measuring the predictive power of one event with respect to another. Reducing uncertainty contributes to efficient coding of sensory information and is thought to be essential for associative learning (see Fiser and Aslin, 2002b). Thus, a learning mechanism that allows detection of conditional probability statistics would support more effective learning, including the prediction of the likelihood of future events relative to co-occurrence frequency.

Fiser and Aslin (2002a) examined whether infants were sensitive to conditional

probability statistics in visual input in addition to co-occurrence frequency. They habituated 9-month-olds to looming multi-element scenes, then tested infants' preference for various element pairs that had occurred in the scenes. The researchers found that infants preferred not only element pairs that co-occurred more frequently as embedded elements in scenes, but also pairs that had higher conditional probability (viz., predictability) between elements in the pair. Thus, infants were sensitive to the statistical coherence of the elements within visual scenes in addition to co-occurrence frequency. In sum, this research demonstrates infants' sensitivity to conditional relations in both auditory and visual input, suggesting that statistical learning is a domain-general process. In the next section, we outline research with adults that provides even stronger support for this idea by examining statistical learning of more complex visual stimuli and the generalizability of statistical learning across contexts.

Adults

Although research with infants has begun to demonstrate the robustness of statistical learning for detecting statistical structure in visual scenes and sequences, the complexity of the visual structures examined in infant studies are rather simplistic compared to those examined in studies with adults. For example, research with adults has examined learners' sensitivity to first- as well as higher-order statistics, and has employed more complex multi-element scenes and sequences than those used with infants to examine the flexibility of the representations learners extract from such input.

Fiser and Aslin (2001) explored the range of first- and higher-order statistics that adults compute during passive viewing of visual scenes. Participants viewed a total of twelve shapes, which were divided into six base pairs. Three of these pairs appeared at a time in various positions within either a 3 X 3 or 5 X 5 grid "scene." The relations between any two shapes in a

scene could be described in terms of co-occurrence and conditional probabilities. Each base pair appeared in half of the scenes, such that the probability of co-occurrence of the two shapes in each of the six base pairs was 0.5. Because the two objects composing each base pair always occurred together within a scene, shapes within base pairs had a conditional probability of 1.0. Fiser and Aslin found that adults detected first-order statistics (single-shape frequency) as well as several higher-order statistics from the scenes. Specifically, participants detected absolute shape-position relations within the grid and shape-pair arrangements independent of grid position. Most importantly, even when the probabilities of co-occurrence of some base pairs and non-base pairs were equated, adults were still able to distinguish the familiar base pairs based solely on their (higher) conditional probabilities.

The finding that adults are capable of implicitly extracting higher order statistics from static spatially presented visual stimuli led Fiser and Aslin (2002b) to probe this ability further with temporally presented stimuli. In this experiment, adult participants viewed twelve shapes organized into four temporal triplets, such that after the first element of the triplet appeared on the screen, the second and then the third elements of the triplet always followed. There were no pauses or breaks between successive shapes such that the triplet structure could only be learned via temporal-order statistics among pairs or triplets of shapes. Just as with spatially presented visual stimuli, participants became sensitive to first-order as well as higher-order statistics in the temporal shape sequences. Participants retained the frequency of individual shapes and distinguished sequences of shapes presented during familiarization from both novel sequences of familiar shapes and sequences of shapes seen during familiarization but presented less frequently. Interestingly, when frequency information and co-occurrence probabilities were equated, adults were still able to distinguish shape sequences based on differing conditional

17

probabilities.

These demonstrations of visual statistical learning with both temporally and spatially presented input raises the question of how such information is represented and whether such representations might generalize to new contexts. Turk-Browne and Scholl (2009) demonstrated that learning of statistical regularities in temporal shape sequences (finding shape "triplets" in a continuous stream of shapes) was expressed in static spatial configurations of these same shape triplets. Similarly, learning of statistically defined spatial configurations (base pairs, as in Fiser & Aslin, 2001) facilitated detection performance in temporal streams (Turk-Browne & Scholl, 2009). Thus, visual statistical learning in adults appears to produce flexible representations that can be generalized to new situations. Such transferability is likely important for visual statistical learning to be practical in ever-changing real-world visual environments.

Constraints on Statistical Learning

The generalizability of statistical learning across tasks and domains raises the important question of whether and what constraints may exist on statistical learning. If one considers the infinite number of possible statistical relations that could be computed at each level of representation, it becomes clear that for statistical learning to be feasible, it must be constrained. What are these constraints?

Types of Input

It is unlikely that all statistical regularities are learned equally well, given the infinite number of possible statistics that could be extracted from the environment. Rather, research suggests that statistical learning mechanisms preferentially track statistical regularities in the types of input that occur most frequently in the natural environment (Conway & Christiansen, 2009; Emberson, Conway, & Christiansen, 2011; Newport & Aslin, 2004).

Spatial versus sequential input. Intuitively, there seem to be structured differences in the organization of auditory and visual information in the natural environment. For instance, auditory information is conveyed temporally whereas visual information is arrayed spatially. Moreover, each sensory modality seems to process particular aspects of environmental input. For instance, a brief snapshot is typically enough time to recognize a complex visual scene whereas at least several seconds are needed to recognize a voice or melody (Conway & Christiansen, 2009). These intuitions are supported by studies of perception and memory suggesting that spatial information weighs most prominently in visual cognition, whereas temporal information weighs most prominently in audition (see Conway & Christiansen, 2009 for a discussion). Such modality differences raise the question of whether statistical learning processes might be constrained to preferentially track statistics in input that accords with the auditory-temporal, visual-spatial structure of the environment.

Conway and colleagues (Conway & Christiansen, 2005, 2009; Emberson et al., 2011) examined how modality differences may constrain implicit statistical learning. For example, Conway and Christiansen (2009) investigated whether vision and audition exhibited different constraints on statistical learning of spatially and temporally structured information. Conway and Christiansen compared learning of one statistically defined structure presented in three different formats: auditory information presented temporally (pure tones of various frequencies presented one at a time through headphones), visual information presented temporally (different colored squares presented one at a time in the center of screen), and visual information presented spatially (the same colored squares presented simultaneously left to right in a horizontal row across the center of the screen). The task was an artificial grammar learning (AGL) task in which adult learners were presented with a set of training sequences that adhered to a specific rule-

governed finite state grammar. After the learning task, learners were presented with a test on classifying novel sequences as being either legal (generated by the same rules as the training sequences) or illegal. The results demonstrated that participants in the visual-spatial condition classified test sequences with a similar degree of accuracy as participants in the auditory condition. However, participants in the visual-temporal condition were significantly less accurate in their classifications compared to those in the auditory condition. This ability to acquire the structure of spatially arrayed visual input as well as temporally structured auditory, but not visual, input suggests that adults' statistical learning may be constrained to preferentially track statistics in inputs that accord with the auditory-temporal, visual-spatial structure of the environment.

Presentation rate. Of course, human learners, including young infants, provide evidence of detecting statistical patterns in sequential visual input under some circumstances (e.g. Bulf et al., 2011; Fiser & Aslin, 2002b; Kirkham et al., 2002). A recent study by Emberson et al. (2011) helped to reconcile these seemingly contradictory findings by investigating the mediating role of presentation timing in statistical learning of auditory and visual information. Their results suggest that there is an interaction of presentation format (spatial versus sequential) and presentation timing in constraining statistical learning across modalities.

Emberson et al. (2011) compared visual and auditory statistical learning in an interleaved familiarization design. Adult learners were presented with a visual stream of abstract shapes organized into triplets that was interleaved pseudo-randomly with an auditory stream of monosyllabic nonsense words also organized into triplets. Participants were randomly assigned to either attend to the visual stream or the auditory stream, and given a cover task (detecting repeat elements in only that stream) to ensure that attention was allocated to the appropriate

stream. Following familiarization, participants were tested on learning in each modality. During test trials, participants judged which of two sequences seemed more familiar: a triplet from familiarization or a foil sequence that did not adhere to the triplet structure. Importantly, this study compared effects of variation in presentation rate. In the "fast" condition, elements were presented for 225 ms with an ISI of 150 ms, resulting in an SOA of 375 ms. In the "slow" condition, elements were presented with an SOA of 750 ms.

Emberson et al. (2011) found that performance in the unattended modality did not differ from chance in any condition. At the fast presentation rate, the statistical relations between adjacent elements were only learned in the attended *auditory* stream. At the slow presentation rate, the opposite effect occurred: only the relations between adjacent elements in the attended *visual* stream were learned. Emberson et al. posited that visual statistical learning improved with the slower rate of presentation because it was less temporally demanding on the visual system. In contrast, auditory statistical learning was impaired at the slower presentation rate because of weaker perceptual grouping cues. That is, when sequential elements were separated by longer intervals, they were less likely to form a single perceptual unit or stream, hindering the detection of statistical information in the stream. Taken together, these results document complex constraints on statistical learning that accord with the structure of the natural environment, with relatively rapid presentation of temporal information critical for auditory statistical learning, and either static spatial information or relatively slowly presented temporal information critical for visual statistical learning.

Natural language: types of non-adjacent regularities. This interaction of presentation format and timing in statistical learning illustrates one way in which constraints on the types of information over which statistical learning operates may reflect environmental structure. Some

researchers have additionally argued that constraints on learning not only reflect, but also help to explain, structural aspects of the environment, such as those found in natural languages (e.g., Christiansen & Chater, 2008). For example, a wide range of adjacent regularities appear throughout natural languages, but the types of non-adjacent regularities languages exhibit are quite constrained.

Newport and Aslin (2004) investigated the intriguing possibility that constraints on the types of non-adjacent statistical computations that learners perform may match and even drive observed constraints on non-adjacent regularities in natural languages. For example, it is common for natural languages to contain non-adjacent regularities relating elements of one kind while skipping over intervening elements of a different kind. In Hebrew and Arabic, word stems are formed out of phonemic segments of one kind (consonants), while intervening segments are of another kind (vowels). In contrast, it is uncommon for natural languages to contain non-adjacent regularities in which intervening items are of the same kind as that in which the non-adjacent regularities occur. Newport and Aslin examined adults' detection of conditional relations among non-adjacent elements that did and did not adhere to this natural language structure: non-adjacent consonants (with one unrelated intervening vocalic segment), non-adjacent vowels (with one unrelated intervening consonantal segment), and non-adjacent syllables (with one intervening syllable that was unrelated). In accord with the structure of natural languages, adults seemed to be unable to track the relations between non-adjacent syllables, where the intervening element was of the same kind (a syllable). Even when the patterns were quite simple and participants were given extensive exposure to the patterns (in one case over 10 days of repeated exposures), participants remained unable to track relations between non-adjacent syllables. In contrast, adults readily learned the relations between non-adjacent

consonants and vowels, where the intervening element was a different kind from that in which

the non-adjacent regularities occurred. These findings suggest that constraints on statistical

learning may help to explain the universal aspects of these patterns in natural languages. Similar

to Conway and colleagues' results (Conway & Christiansen, 2009; Emberson et al., 2011), these

findings also demonstrate that human learners preferentially track statistical information only in

particular types of environmental input. Such findings highlight the importance of considering

statistical learning in its broader environmental context, including the nature of the input to

which the learner is exposed, as well as the cognitive, developmental, and attentional state of the

learner.

The State of the Learner

Human learners are characterized by perceptual biases and cognitive constraints.

Appreciating the influences of learners' biases and developmental state on statistical learning is

necessary for a complete understanding of the extent and limits of this domain-general learning

process across development.

Spatiotemporal biases and perceptual similarity. Consideration of learners' perceptual

biases is especially important for understanding constraints on visual statistical learning, as such

biases have been shown to influence the types of statistics learners extract from visual scenes

(Fiser, Scholl, & Aslin, 2007). One general perceptual bias exhibited by infants and adults is the

bias to perceive objects as moving along specific trajectories given certain visual and/or auditory

cues (e.g. Sekuler & Sekuler, 1999; Shimojo, Watanabe & Scheier, 2001). When observing two

identical objects moving toward each other, coinciding, then moving away from each other, two

interpretations are possible: (1) the two objects streamed past one another (*streaming*), or (2), the

two objects bounced off of one another (*bouncing*). Various perceptual features such as the

acceleration of the objects (Fiser, et al., 2007; Sekuler & Sekuler, 1999) or the presence of a sound at the time of coincidence (Sekuler, Sekuler, & Lau, 1997; Watanabe & Shimojo, 2001) bias observers toward one of these two interpretations.

Fiser et al. (2007) investigated whether this perceptual bias to perceive objects as moving along specific trajectories affected the types of statistics adult learners computed from visual events. Participants observed a single object move behind an occluder and then saw two objects emerge from behind the occluder simultaneously. One object emerged from the occluder following the same trajectory as the first object. The second object emerged from the occluder at a 90° angle to the original trajectory. Thus, presentations could be interpreted two different ways: (1) as an object streaming behind the occluder on a straight trajectory, or (2) as an object bouncing off of a surface behind the occluder and reemerging on the same side that it originated.

To examine whether perceived motion trajectories would bias statistical learning, Fiser et al. (2007) manipulated the acceleration of the objects to bias observers toward one of these two percepts. Objects moving at constant speed produced a streaming percept whereas decelerating-accelerating objects produced a bouncing percept. If visual statistical learning mechanisms compute all available temporal co-occurrences of shape pairs, then learners should acquire transitions from the first shape to each of the two later shapes equally well, regardless of whether observers were biased toward streaming or bouncing percepts. However, this is not what Fiser et al. found. Rather, adults preferentially learned the associations consistent with the perceptual bias of streaming or bouncing they had during familiarization. Thus, this perceptual bias constrained statistical learning to shape pairs consistent with that bias.

The influence of perceptual biases on statistical computations is not limited to statistics in visual scenes. Similar to spatiotemporal biases, Gestalt principles of perception have been shown

to constrain the detection of statistical relations in both auditory and visual input (Baker, Olson & Behrmann, 2004; Creel, Newport & Aslin, 2004; Emberson et al., 2011; Newport & Aslin, 2004). For example, Creel et al. (2004) demonstrated that Gestalt principles of element similarity interact with temporal adjacency in determining what kinds of auditory statistical regularities are learned. In this experiment, adult participants were presented with two interleaved streams of tone triplets such that participants heard the first tone of the first triplet stream, followed by the first tone of the second triplet stream, then the second tone of the first stream, then the second tone of the second stream, and so on (Creel et al., 2004). The result of this interleaving was that triplets could only be detected via sensitivity to non-adjacent conditional relations.

Interestingly, adults showed no learning of the tone triplets, only sensitivity to the less reliable relations between adjacent elements in the stream. However, when Creel et al. (2004) included perceptual grouping cues, by presenting the two interleaved streams in differing pitch ranges or timbres, adults became sensitive to the conditional relations between the similar, yet temporally non-adjacent, elements. This finding suggests that Gestalt principles of similarity interact with temporal adjacency in constraining statistical learning.

Availability of cognitive resources. Thus far, our discussion has highlighted similarities in infants' and adults' sensitivities to statistical information. Researchers hold differing views, however, on how implicit statistical learning abilities may change across development (e.g., Janacsek, Fiser, & Nemeth, 2012; Thomas, Hunt, Vizueta, Sommer, Durston, Yang, & Worden, 2004) or remain constant across development (e.g., Reber, 1993; Vinter & Perruchet, 2000).

In some studies reporting developmental differences, older individuals show better learning than younger individuals (e.g., Maybery, Taylor & O' Brien-Malone, 1995). Consistent with this possibility, infants provide evidence for tracking increasingly complex statistical

regularities in visual sequences with age: 2- 5- and 8-month-old infants distinguished structured from random sequences composed of six looming shapes (Kirkham et al., 2002), but newborn infants seemed to distinguish structured from random structure in sequences containing four, but not six, items (Bulf et al., 2011).

In other cases, however, younger individuals outperform older individuals (e.g., Janacsek et al., 2012; Jost, Conway, Purdy & Hendricks, 2011). Jost et al. (2011) compared the time course of children's and adults' implicit learning by examining participants' ERPs during a visual statistical learning task. Participants observed a series of stimuli presented one at a time on a screen and pressed a button whenever the target stimulus appeared, which was predicted at different levels of probability by the stimuli immediately preceding the target. Jost et al. found that children exhibited learning-related ERP components earlier in the study than adults, suggesting that children required less exposure to the patterns to detect the statistical structure.

To explain differences in statistical learning ability across development, researchers have appealed to domain-general, maturational constraints on perception and memory. Bulf et al. (2011) suggested that newborns' limited attentional and working memory capacities may inhibit statistical learning efficiency. Interestingly, researchers have posited a similar explanation to account for findings that children outperforming adults. In this case, however, researchers have offered the paradoxical idea that maturational constraints on perception and memory confer a computational advantage for some types of learning (e.g., Elman, 1993; Newport, 1988, 1990). In particular, Newport's (1990) "Less is More" hypothesis assumes that children's abilities to perceive and store complex stimuli is reduced compared to those of adults, and suggests that such limitations give children an advantage for tasks requiring componential analysis because

children are better able to identify and process component parts. Adults, in contrast, attempt to perceive and store stimulus relations of greater complexity.

Suggestions that maturational constraints on perception and memory can both hurt and help performance in tasks requiring componential analysis appear contradictory. However, most empirical support for Newport's "Less is More" hypothesis (1990; e.g., Kersten & Earles, 2001) comes from child and adult populations, leaving open the possibility that very early increases in infants' relatively limited perception and memory abilities may be positively related to statistical learning ability. To our knowledge, however, Bulf et al.'s (2011) hypothesis that limited cognitive resources limit newborns' statistical learning performance has not yet been confirmed independently. Visual working memory performance increases roughly linearly across the first postnatal year (Diamond, 1985; see Bell & Morasch, 2007 for a review), and a number of other early developments could, in principle, be responsible for changes in statistical learning (e.g., different spatiotemporal biases due to changes in perceptual acuity). An important avenue for future research will be to investigate these possibilities, beginning by examining the relation between the development of infant working memory ability and statistical learning ability.

In addition to maturational constraints on perception and memory, the allocation of attentional resources may also play a role in constraining statistical learning. Although some researchers have argued that statistical learning is an "automatic" (i.e., implicit, rapid) process (e.g. Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), other researchers have found reason to suggest that statistical learning both is and is not automatic (e.g. Turk-Browne, Junge, & Scholl, 2005). It is automatic in that statistical computations seem to be carried out without conscious intent and often without awareness that any structure was learned (e.g. Meulemans, Van der Linden, & Perruchet, 1998; Saffran et al., 1997; Turk-Browne et al., 2005). However, statistical

27

learning is not automatic in that it operates better over attended versus unattended input (e.g. Emberson et al., 2011; Toro, Sinnett & Soto-Faraco, 2005; Turk-Browne, et al., 2005). For instance, when two interleaved streams of shapes are presented to observers in two different colors, and participants are instructed to attend to only one color, only the statistical relations in the attended color are learned (Turk-Browne, et al., 2005). This attentional constraint on statistical learning appears to be one of its most general limitations, likely constraining detection of statistical regularities regardless of input domain or modality (e.g. Emberson et al., 2011).

Prior experience. In addition to maturational changes in cognitive resources, such as working memory capacity and attention, another important aspect of development is learning from experience interacting with the environment. Expectations about the structure of the environment undergo rapid changes in the first years after birth due to experiences interacting with the world (e.g. Adolph, Eppler, & Gibson, 1993; Campos, Bertenthal, & Kermoian, 1992). Such changes in learners' expectations about the structure of their environment may have the potential to influence statistical learning processes (Thiessen, 2010). For example, years of experience with language may provide adults with strong expectations that words and objects relate to one another (e.g., Namy & Waxman, 1998).

Thiessen (2010) investigated how such expectations influence adults' statistical learning of word-object associations. Adults were presented with paired audio-visual information in which word boundaries as well as word-object associations were statistically defined. Participants tracked both of these statistical relations simultaneously, and word segmentation benefited from the addition of word-object associations. When adults were presented with tonal rather than linguistic stimuli, however, they did not benefit from the regular relations between tone words and objects. Thiessen suggested that experience with language may predispose adults

to expect words and objects to relate to one other, such that they are sensitive to these associations in linguistic input, but not in tonal input. This hypothesis leads to the prediction that young infants may not benefit from word-object relations even with linguistic input, because they may not yet have built up the expectation that words relate to objects (e.g. Werker et al., 1998). This is precisely what Thiessen found; similar to adults in the tonal condition, 8-month-old infants' ability to segment words did not benefit from the presence of word-object relations, regardless of whether linguistic or nonlinguistic input was used.

Thiessen's (2010) findings demonstrate the role of prior experience and learners' expectations in facilitating computation of previously ignored statistics. Other research, however, indicates that prior experience can impede statistical computations. For example, Gebhart, Aslin and Newport (2009a) presented adult learners with auditory sequences of trisyllabic nonsense words defined by the TPs between syllables. When the researchers altered the organization of the nonsense words mid-way through the familiarization stream, participants only learned the first of the two structures. Participants detected words in both structures only when exposure to the second structure was tripled in duration, or when the transition between structures was explicitly marked. Thus, successful extraction of the statistical regularities in one auditory structure inhibited learning of a subsequent auditory structure.

Mechanisms Underlying Statistical Learning

How is it that statistical learning can be so constrained while still adapting flexibility to input across domains and modalities? The reason for both flexibility and constraints on statistical learning is likely because the environment contains both variance and invariance; organisms need a way to flexibly adapt and generalize to different contexts while simultaneously honing in on the types of structures that are most consistent and informative in the environment. What is

less clear are the mechanisms by which statistical learning occurs and how these mechanisms are configured to allow for both flexibility and constraints.

We began this review by introducing statistical learning as sensitivity to transitional probabilities (TPs), and this view was predominant in the early days of infant statistical learning research that focused predominantly on word segmentation. However, there is now a wealth of data on infants' and adults' statistical learning across domains, and this calls for a broader view of statistical learning (e.g., Frank et al., 2010; Graf Estes et al., 2007; Maye, Werker, & Gerken, 2002; Saffran, 2001; Smith & Yu, 2008; Thiessen & Saffran, 2003). For example, consider Saffran's (2001) and Graf Estes et al.'s (2007) findings that the output of statistical learning is entire word-like units, not simply highly probable sound sequences. A mechanism that only tracks probabilistic relations between elements cannot fully account for such a finding (see Thiessen, Kronstein, & Hufnagle, in press). Moreover, even in segmentation tasks, models designed to track transitional probabilities do not always accord well with human performance (see Frank et al., 2010).

A variety of alternate models of statistical learning have been proposed that do not rely on explicitly computed statistics. It is not yet clear which type of model produces the most valid account of human learning processes across tasks (Frank et al, 2010). A complete review of all such models is beyond the scope of this review; instead, we briefly describe one well-known model, PARSER (Perruchet & Vinter, 1998), to illustrate that there are multiple possible mechanisms to account for statistical learning data.

PARSER (Perruchet & Vinter, 1998) is a type of "chunking" model that produces the same segmentation results as Saffran et al. (1996a, 1996b) by implementing basic laws of attention, memory, and associative learning, rather than by computing statistics such as

transitional probabilities. PARSER is modeled on the principle that perception guides internal representation. Briefly, units that are perceived within one attentional focus are "chunked" into a new representational unit. The fate of these new representations depends on fundamental principles of memory: Internal representations of chunks that are repeated are progressively strengthened, and representations of chunks that are not repeated are forgotten (Perruchet & Vinter, 1998). Applied to Saffran et al.'s (1996a, 1996b) segmentation task, PARSER would first randomly segment the speech stream into small chunks. Because chunks have a greater chance of being repeated if they are part of the same word than if they span a word boundary, internal representations of words or parts of words will be stronger in memory than representations of non-words and chunks spanning word boundaries. Thus, PARSER can account for Saffran et al.'s (1996a, 1996b) findings of participants' greater sense of familiarity for words than non-words or part-words.

As noted, several models of statistical learning employing quite different mechanisms have been proposed to account for the various findings of the statistical learning literature, but no model has yet been proposed that can account well for human performance across statistical learning tasks (Thiessen et al., in press). In particular, what is lacking are models that achieve sensitivity to other statistical relations in addition to conditional relations, such as the central tendency of a set of elements (distributional statistical learning; e.g., Maye, Werker, & Gerken, 2002), as well as models that account for human's learning and generalization based upon similarity across items extracted from the input (e.g. Thiessen & Saffran, 2003). Thiessen et al. (in press) argued that mechanisms designed only to account for the extraction of units, such as segmenting words from a speech stream, cannot account for these other forms of statistical learning.

Thiessen et al. (in press) proposed a framework that attempts to account for these various forms of statistical learning by combining processes of extraction with processes of comparison across extracted segments in an iterative model whereby the discovery of new structures via comparison serves to educate the extraction processes. To illustrate this idea, consider the finding that when syllable stress and statistical cues indicated different word boundaries in a speech stream, 7-month-olds segmented based on statistical cues, whereas 9-month-olds segmented based on stress cues (Thiessen & Saffran, 2003). Models that are only designed to account for segmentation cannot explain these findings without positing additional changing constraints on the learner or on the statistical learning mechanism itself. In contrast, Thiessen et al.'s (in press) framework accounts for such findings without necessitating new or changing constraints; according to this framework, such findings demonstrate initial segmentation based on conditional statistics followed by comparison across segmented words, allowing the discovery of patterns of stress cues in English words, which in turn inform the process of segmentation in the future.

Although Thiessen et al.'s (in press) framework has not yet been implemented into a working computational model, such a framework pushes the field forward by offering a mechanism that accounts for developmental differences in statistical learning. Moreover, this framework is also helpful for thinking about the origins of the constraints on and flexibility of statistical learning. That is, the framework is based on general processes of attention, memory, and comparison that likely govern extraction and generalization across domains. Furthermore, this framework describes a way in which learners may use a constrained, limited-capacity mechanism to flexibly adapt to different characteristics of the input over time.

Conclusion

Statistical learning is a means of uncovering structure in complex environmental input. It operates in both auditory and visual domains, and encodes multiple types of statistics simultaneously. Constraints on statistical learning serve to reduce the number of possible associations available, making statistical learning tractable.

A comprehensive model of statistical learning across domains has not yet been reported in the literature, but much progress has been made in uncovering the origins of both the flexibility of and constraints on statistical learning. Specifically, flexibility may be the result of mechanisms built upon domain-general processes, such as attention, memory, and perception, rather than domain- or modality-specific processes. Flexibility may be built into the system as a product of learners' ability to discover new structures via comparison, and use those new structures to influence further extraction (Thiessen et al., in press). Constraints on statistical learning are driven by a variety of factors: limited attention, perception, and memory capacity, as well as maturational increases in these domain-general processes; learned biases and expectations about the structure of the environment; and ways in which statistical tendencies in language have been shaped to fit the human brain, rather than vice versa.

Thus, while research has revealed numerous influences on the various constraints on statistical learning, the principal contribution to flexibility in statistical learning appears to be its domain-general nature. Nevertheless, the domain-generality of statistical learning mechanisms has been hotly debated. Some researchers interpret demonstrations of statistical learning across domains and modalities as evidence of a single, domain-general statistical learning mechanism (e.g. Kirkham et al., 2002), but others contend that statistical learning cannot be domain general due to observed modality-specific constraints (Conway & Christiansen, 2005, 2009; Emberson et

al., 2011). Specifically, they cite findings such as the auditory-temporal, visual-spatial distinction as evidence for separate statistical learning mechanisms for each modality (Conway & Christiansen, 2009). One limitation of this line of reasoning, however, is that constraints differentially affecting statistical learning of different types of input *within* modalities (e.g. Endress 2010; Thiessen, 2010) would necessitate multiple statistical learning mechanisms *within* modalities as well as across modalities. Thus, the domain-general view seems to be the most parsimonious account of the data. However, evidence supporting a domain general account of statistical learning does not exclude the possibility of multiple domain- or modality-specific statistical learning subsystems. Further research is needed to determine which of these views provides the most complete account of statistical learning. Research examining statistical learning performance using comparable tasks across domains and modalities, as well as research comparing the ability of modality-specific and domain-general computational models to fit such human data, may be particularly informative.

Moreover, future research should continue to investigate the type of flexibility in statistical learning documented by Turk-Browne and Scholl (2009), who demonstrated flexibility in the transferability of the representations that emerged from adults' visual statistical learning. Further research should pursue similar lines of research employing other tasks and input types to investigate the generalizability of such findings across modalities. A final important avenue for future research will be to continue working towards developing a comprehensive model that can accommodate the various forms of statistical learning (sensitivity to conditional relations, distributional statistics) across domains as well as developmental changes in such learning. Longitudinal research and research that makes within-subjects comparisons across tasks may be particularly useful in this endeavor.

References

Adolph, K. E., Eppler, M. A., and Gibson, E. J. (1993). Crawling versus walking infants' perception of affordances for locomotion over sloping surfaces. *Child Development* 64, 1158-1174.

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

Baker, C. I., Olson, C. R. and Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science* 15, 460-466.

Bell, M. A., and Morasch, K. C. (2007). "The development of working memory in the first 2 years of life," in *Short and Long-term Memory in Infancy and Early Childhood,* eds. L. M. Oakes and P. J. Bauer (New York: Oxford University Press), 27-50.

Bulf, H., Johnson, S.P., and Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition* 121, 127-132.

Brady, T.F., and Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science* 19, 678-685.

Campos, J. J., Bertenthal, B. I., and Kermoian, R. (1992). Early experience and emotional development: The emergence of wariness of heights. *Psychological Science* 3, 61-64.

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences,* 31, 489–509.

Christiansen, M. H., Conway, C. M., and Onnis, L. (2007). "Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms," in *Proceedings of the 29th annual meeting of the cognitive science society, eds* D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 173–178.

Conway, C.M., and Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, 24–39.

Conway, C.M. and Christiansen, M.H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology* 21, 561–580.

Creel, S.C., Newport, E.L., and Aslin, R.N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology* 30, 1119–1130.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development* 56, 868-883.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition,* 48, 71-99.

Emberson, L.L., Conway, C.M., and Christiansen, M.H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology* 64, 1021–1040.

Endress, A.D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica* 135, 182-190.

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

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

Fiser, J., and Aslin, R. N. (2002b). Statistical learning of higher-order temporal structure from

visual shape sequences. *Journal of Experimental Psychology: Learning Memory and Cognition* 28, 458-467.

Fiser, J., and Aslin, R. N. (2005). Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General* 134, 521-537.

Fiser, J., Scholl, B. J., and Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin and Review* 14, 173 - 178.

Frank, M.C., Goldwater, S., Griffiths, T., and Tenenbaum, J.B. (2010). Modeling human performance in statistical word segmentation. *Cognition,* 117, 107-125.

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

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009b). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin and Review* 16, 486-490.

Graf Estes, K. M., Evans, J., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science* 18, 254-260.

Harris, Z. (1955). From phoneme to morpheme. *Language,* 31, 190-222.

Hayes, J. R., and Clark, H. H. (1970*). "*Experiments in the segmentation of an artificial speech analog" in *Cognition and the development of language*, ed. J.R. Hayes (New York: Wiley), 221-234.

Hunt, R.H., and Aslin, R.N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General* 130*,* 658-680.

Janacsek, K., Fiser, J., and Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science, 15*, 496-505.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339-345.

Jost, E., Conway, C. M., Purdy, J. D., and Hendricks, M. A. (2011). Neurophysiological correlates of visual statistical learning in adults and children. Paper Presented at the 33rd Annual meeting of the Cognitive Science Society, July 2011.

Kersten, A. W. and Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language, 44*, 25–273.

Kirkham, N.Z., Slemmer, J.A., and Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition* 83, B35–B42.

Kirkham, N.Z., Slemmer, J.A., Richardson, D.C., and Johnson, S.P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development* 78, 1559-1571.

Lany, J., and Gomez, R.L. (2008). 12-Month-Olds benefit from prior experience in statistical learning. *Psychological Science* 19, 1247-1252.

Maye, J., Werker, J. F. and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Maybery, M., Taylor, M., & O'Brien-Malone, A. (1995). Implicit learning: sensitive to age but not IQ. *Australian Journal of Psychology, 47*, 8–17.

Meulemans, T., Van der Linden, M., and Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology* 69, 199–221.

Namy, L. L., and Waxman, S. R. (1998). Words and gestures: Infants' interpretation of different

forms of symbolic reference. *Child Development* 69, 295–308.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of

the acquisition of American Sign Language. *Language Sciences,* 10, 147–172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science,* 14, 11–

28.

Newport, E. L., and Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of

nonadjacent dependencies. *Cognitive Psychology* 48, 127–162.

Pelucchi, B., Hay, J. F. and Saffran, J. R. (2009). Statistical learning in a natural language by 8-

month-old infants. *Child Development* 80, 674–685.

Perruchet, P. and Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of

Memory and Language,* 39, 246-263.

Reber, A.R. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive

unconscious.* New York: Oxford University Press.

Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition* 81,

149–169.

Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and

Language* 47, 172–196.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current

Directions in Psychological Science* 12, 110-114.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old

infants. *Science* 274, 1926–1928.

Saffran, J.R. and 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., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

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

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science* 8, 101–105.

Saffran, J. R. and Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy* 4, 273–284.

Sekuler, A. B. and Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception* 28, 415-432.

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308.

Shimojo, S., Watanabe, K., and Scheier, C. (2001). "The resolution of ambiguous motion: Attentional modulation and development," in *Visual Attention and Cortical Circuits,* eds. J. Braun, C. Koch, and J. Davis (Cambridge, Massachusetts: MIT Press), 243–264.

Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition,* 106, 1558-1568.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology,* 50, 86–132.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*

10, 21.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science* 34, 1093–1106.

Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development* 82, 462–470.

Thiessen, E.D., Hill, E. E., and Saffran, J.R. (2005). Infant-directed speech facilitates word segmentation. *Infancy,* 7, 53–71.

Thiessen, E.D., Kronstein, A.T., and Hufnagle, D.G. (in press). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin.*

Thiessen, E. D. and Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology,* 39, 706-716.

Thiessen, E. D. and Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development,* 3, 73-100.

Thomas, K.M., Hunt, R.H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., and Worden, M.S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of Cognitive Neuroscience,* 16, 1339–1351.

Toro, J. M., Sinnett, S., and Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition* 97*,* B25-B34.

Turk-Browne, N. B., and Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance* 35, 195 - 202.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., and Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

34, 399 - 407.

Turk-Browne, N. B., Junge, J. A., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General* 134, 552 - 564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., and Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience* 21, 1934 - 1945.

Vinter, A., and Perruchet, P. (2000). Implicit learning in children is not related to age: Evidence from drawing behavior. *Child Development,* 71, 1223–1240.

Watanabe, K. and Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science* 12, 109-116.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology* 34, 1289– 1309.

PAPER 2:

Modality constraints on statistical learning in infancy

Abstract

Research has demonstrated infants' ability to detect statistical relations among auditory and visual inputs. It remains unclear, however, how auditory and visual statistical learning (SL) ability compare in infancy. Four experiments with 71 8-month-olds examined the limits of infants' visual SL (VSL) ability. In contrast to previous auditory findings, infants were not sensitive to transitional probability statistics in visual triplet sequences (Experiments 1-2), though infants discriminated random from structured visual sequences (Experiment 3). Additionally, individual differences in VSL performances related to cognitive ability, but not visual short-term memory (Experiment 4). Detecting transitional probabilities in visual sequences appears challenging for infants. Nevertheless, individual differences in VSL may be meaningful in terms of both VSL ability and early cognitive development more broadly.

Modality Constraints on Statistical Learning in Infancy

Every day, learners encounter an abundance of complex auditory and visual information. How is it that learners transform these sensory inputs into mental representations? One step in this process likely involves detecting statistical regularities within the input. Research has demonstrated powerful statistical learning (SL) abilities in adult and infant learners, including the ability to extract statistical relations between auditory inputs such as syllables in a sequence (e.g., Pelucchi, Hay, & Saffran, 2009b; Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Thiessen, Hill, & Saffran, 2005) and between visual inputs such as spatially or temporally presented shape patterns (e.g., Bulf, Johnson, & Valenza, 2011; Fiser & Aslin, 2001, 2002a,b; Kirkham, Slemmer, & Johnson, 2002; Kirkham, Slemmer, Richardson, & Johnson, 2007; Marcovitch & Lewkowicz, 2009; Slone & Johnson, 2015). In adults, input modality (e.g., visual, auditory) can affect the extent to which statistical regularities are learned (Conway & Christiansen, 2005, 2009; Emberson, Conway, & Christiansen, 2011; Saffran, 2002). Moreover, a number of studies suggest that SL performance is superior for auditory compared to visual information in adults (Conway & Christiansen, 2005, 2009; Robinson & Sloutsky, 2007). While these studies suggest that SL is not an amodal, abstract learning mechanism, it remains unclear how auditory and visual SL ability compare in infancy. Given that SL ability has been posited to underlie cognitive skills ranging from infants' language acquisition to apprehension of event structure (e.g., Bates & Elman, 1996; Roseberry, Richie, Hirsh-Pasek, Golinkoff, & Shipley, 2011), understanding the limits/extent of SL ability across modalities in infancy is critical.

A number of studies have demonstrated infants' SL ability using auditory (e.g., Saffran, Johnson, Aslin & Newport, 1999) or visual (e.g., Fiser & Aslin, 2002b; Kirkham et al., 2002) stimuli. This has led some researchers to argue that infants' SL is domain-general, operating

similarly across perceptual modalities (Kirkham et al., 2002). However, infant auditory SL

(ASL) and visual SL (VSL) studies have examined different statistical structures and tested

statistical learning differently, making it difficult to compare learning outcomes across

modalities. For instance, ASL studies typically employ more complex statistical structures and

more rigorous test comparisons than VSL studies. The present series of experiments was

therefore designed to address two related questions: (1) Can infants acquire via VSL the same

complexity of statistical structures that infants acquire via ASL? (2) What factors might

constrain VSL in infancy?

Infants' Auditory Statistical Learning

SL was first proposed as a mechanism to facilitate language learning. In a seminal study,

Saffran et al. (1996a) showed that infants could use transitional probability statistics to discover

word boundaries in fluent speech. Transitional probability (TP) is defined as the probability of

event Y given event X, and is a measure of the strength with which X predicts Y. Saffran et al.

hypothesized that learners could track TPs between adjacent syllables, using peaks in TP to

group syllables into words, and dips in TP to identify breaks between words.

To test this hypothesis, Saffran et al. (1996) constructed an artificial speech stream by

pseudo-randomly concatenating four three-syllable "words." There were no breaks or pauses

between words such that the speech stream could only be segmented into its component words

by tracking differences in TPs between syllables. After 2 minutes of exposure to this speech

stream, 8-month-olds were presented with 12 test trials. Each test trial consisted of one 3-syllable

sequence, presented repeatedly with 500ms pauses between repetitions, until the infant

disengaged attention. Half of the test trials presented a word from the speech stream (high TP

between syllables), and the other half of test trials presented a syllable sequence that had spanned

word boundaries – a "part-word" (lower TP between syllables). Infants listened significantly longer to part-words than to words, demonstrating sensitivity to the statistical structure of the input, and suggesting SL as a possible mechanism contributing to language acquisition.

The nearly two decades following this landmark study have yielded an abundance of research on infant SL with a variety of auditory inputs. By 8 months of age, infants are sensitive to TPs between items in artificial languages (e.g., Aslin, Saffran, & Newport, 1998), unfamiliar natural languages (e.g., Pelucchi, Hay, & Saffran, 2009b), and non-linguistic auditory sequences (Saffran et al., 1999). Moreover, infants can track multiple levels of statistical structure simultaneously (Saffran & Wilson, 2003), and treat the output of SL as actual linguistic items (Saffran, 2001), mapping them as labels for objects (e.g., Graf Estes, Evans, Alibali, & Saffran, 2007).

Infants' Visual Statistical Learning

Infants' ability to detect a variety of statistical regularities from both linguistic and non-linguistic auditory materials after only brief exposure prompted researchers to begin examining the modality-specificity of SL. These investigations have primarily focused on infants' acquisition of visual sequence structures (cf. Fiser & Aslin, 2002b).

Kirkham et al. (2002) were the first to examine infants' SL of visual sequences. Infants viewed a continuous stream of six colored shapes, organized into three pairs but presented one at a time, on a computer monitor. Similar to the Saffran et al. (1996a) auditory sequences, the visual sequence was created by pseudo-randomly concatenating shape pairs such that pair boundaries were marked solely by differing TPs within units (1.0 TP) and between units (.33 TP). After habituating to this sequence, infants saw six test trials, each a continuous sequence of shapes. In half of the test trials the sequence followed the pair structure of the habituation

47

sequence. The other half of test trials presented a randomly ordered sequence of the same six

shapes.. Infants at 2, 5, and 8 months looked longer to the random sequence than to the familiar

habituation sequence. The authors interpreted this result as evidence of a domain-general SL

mechanism, unconstrained by modality.

More recent investigations of infants' VSL employed stimuli and methods similar to

those of Kirkham et al. (2002): habituation sequences consisting of continuous streams of six or

fewer shapes organized into pairs (Bulf et al., 2011; Kirkham, et al. 2007; Marcovitch &

Lewkowicz, 2009; Slone & Johnson, 2015), and post-habituation test sequences consisting of

ordered sequences of shape pairs versus randomly-ordered sequences (Bulf et al., 2011; Kirkham

et al., 2007). These studies demonstrated infants' initial capacity for detecting statistical

regularities (at some level) in the visual modality, but it remains unclear from these experiments

whether infants can acquire via VSL the same complexity of statistical structures that infants

acquire via ASL.

Comparing ASL and VSL Tasks for Infants

Infant ASL and VSL studies differ in at least two crucial ways. First, ASL studies

typically employ more complex stimuli than VSL studies. For instance, whereas ASL studies

often examine sequences of 10 to 12 items organized into triplets (e.g., Saffran et al., 1996a) or a

combination of pairs and triplets (e.g., Thiessen et al., 2005), VSL studies typically use

sequences of six or fewer items organized into pairs (Bulf et al., 2011; Kirkham et al., 2002,

2007; Marcovitch & Lewkowicz, 2009; Slone & Johnson, 2015). Second ASL studies typically

test infants' ability to discriminate high- from low-probability test items, such that sensitivity to

specific statistical regularities such as TP can be determined (e.g., Saffran, Aslin, & Newport,

1996; Thiessen et al., 2005; Saffran et al., 1999). In contrast, VSL studies typically examine

whether infants can discriminate a familiarized structured shape sequence from a randomly-ordered sequence (Bulf et al., 2011; Kirkham et al., 2002, 2007) and as such cannot specify exactly what statistical information infants tracked during learning trials (see Slone & Johnson, 2015 for a discussion). Adding to this concern, a recent study demonstrated spontaneous preferences for random versus structured shape sequences in 5-month-olds (Addyman & Mareschal, 2013) suggesting that preferences for the random test sequence in VSL studies may not in fact have resulted from SL.

The Current Study

The present series of experiments was therefore designed to address two related questions: (1) can infants acquire via VSL the same complexity of statistical structures that infants acquire via ASL? and (2) what factors might constrain VSL in infancy? We presented infants with sequences composed of triplets of items rather than pairs of items, and examined whether infants showed evidence of VSL when required to discriminate high- from low-TP test items (Experiments 1-2), or whether infants only show evidence of learning when asked to discriminate random from structured sequences (Experiment 3). Experiment 4 addressed the second question by examining the role of several factors – visual short-term memory, and verbal and nonverbal cognitive ability – in constraining infants' VSL.

<div align="center">

Experiment 1

</div>

To examine whether infants can acquire via VSL the same complexity of statistical structures that infants acquire via ASL, Experiment 1 employed a method of testing VSL that presented familiarization and test stimuli similar in complexity to those described by Saffran et al. (1996a). Specifically, 12 shapes organized into four triplets served as the training stimulus,

followed by test sequences designed to examine infants' ability to discriminate high-TP shape sequences from low-TP shape sequences.

Half the infants participated in a "3-1-2" condition in which they were tested for sensitivity to TPs in 1-2-3 (a triplet from the learning sequence) versus 3-1-2 (the last item of one triplet followed by the first two items of a different triplet) test sequences (cf. Saffran et al., 1996a). Successful discrimination of these test sequences requires that infants note differences in TPs between items at the beginnings of test units – the low-TP transition from an item 3 to an item 1, compared to the high-TP transition from an item 1 to an item 2. In this respect, infants are tested for attention to primacy. The other half of infants participated in a "2-3-1" condition in which they were tested for attention to recency. We included this second test condition because VSL studies with adults have yielded conflicting results as to whether learners show differential sensitivity to statistical information at the beginning (a primacy effect) versus the end (a recency effect) of sequences (e.g., Conway & Christiansen, 2005, 2009; Fiser & Aslin, 2002a). Infants in the 2-3-1 condition were tested for sensitivity to TPs in 1-2-3 versus 2-3-1 (the last two items of one triplet followed by the first item of a different triplet) test sequences. Successful discrimination of these test sequences requires that infants note differences in TPs between items at the ends of test units. Examining performance on both 3-1-2 and 2-3-1 test sequences ensured detection of VSL even if infants preferentially tracked statistical regularities at only the beginning or the end of triplets. Moreover, comparison of 3-1-2 and 2-3-1 performances allowed us to examine possible recency or primacy effects in infants' VSL.

Method

Participants. Thirty-two healthy full-term 8-month-olds were tested using an infant-controlled habituation technique. Participants were assigned to participate in either a 3-1-2 ($N =$

16; 9 females; *M* age 8 months 2 days, range = 7;14 to 8;18) or 2-3-1 (*N* = 16; 11 females; *M* age

7 months 24 days, range = 7;15 to 8;0) test condition. Data from an additional 12 infants were

excluded from the final sample due to failure to habituate (*n* = 8), fussiness (*n* = 3), or

experimenter error (*n* = 1). Infants were recruited by letter and telephone from hospital records

and given a small gift (a toy or baby t-shirt) for their participation.

Apparatus and stimuli. A Dell Optiplex 755 computer and 61.5-cm color monitor were

used to present stimuli and collect looking time data. The computer calculated the habituation

criterion online for each infant, and changed displays after the habituation criterion was met. The

primary experimenter observed the infant via a video feed from a camera placed directly below

the monitor and coded looking time behavior online by pressing and releasing a preset keyboard

key. The experimenter was unaware of the stimulus sequence viewed by the infant. Each

habituation and test sequence was presented until the infant visually fixated away from the

monitor for over 2 s or until 90 s of looking had accumulated. Prior to the first habituation trial

and prior to every subsequent trial an attention-getter was shown to attract infants' gaze back to

the screen, whereupon the next trial was started immediately.

Stimuli consisted of twelve colored shapes (blue square, brown hourglass, cyan banner,

green plus, magenta arrow, olive star, orange diamond, pink rainbow, purple octagon, red circle,

white heart, yellow triangle) presented one at a time for 750 ms each. Shapes loomed from 1.5 to

5.5 cm in height (about 1.43°-5.25°) in the center of the screen against a black background. The

habituation sequence consisted of a continuous sequence of four shape triplets, randomly

selected for each infant (e.g. triplet 1: olive star, orange diamond, blue square; triplet 2: brown

hourglass, red circle, white heart; triplet 3: green plus, magenta arrow, yellow triangle; triplet 4:

cyan banner, purple octagon, pink rainbow; Figure 2-1A). Triplets could not repeat and there

were no breaks or delays between shapes or triplets such that the only cues to sequence structure were inter-stimulus TPs (1.0 within triplets, .33 between triplets) and co-occurrence frequencies (8.33% within triplets, 2.78% between triplets).

Procedure. Infants sat on a caretakers' lap approximately 60 cm from the computer monitor. Caretakers were instructed not to interact with the infant or attend to the monitor. The habituation sequence was presented until infants exhibited habituation of looking or until 12 trials elapsed. The habituation criterion was met when looking time to a block of four trials reduced below 50% of the infant's looking time to the first four trials.

Following habituation, infants were presented with three repetitions of two unique test trials in alternation, for a total of six test trials. Each test trial consisted of a repetition of three shapes separated by a 750 ms pause to mark unit boundaries. Half of the infants ($N = 16$) saw repetitions of familiar and 3-1-2 test sequences, and the other half saw repetitions of familiar and 2-3-1 test sequences. The familiar sequence was a triplet from habituation (TP = 1.0 and co-occurrence frequency = 8.33% between the first and second shapes and between the second and third shapes; Figure 2-1B), whereas the 3-1-2 and 2-3-1 "part-sequences" spanned a triplet boundary (3-1-2: the third shape of a habituation triplet, followed by the first two shapes of a different habituation triplet such that the TP and co-occurrence frequency between the first and second shapes of this test sequence were .33 and 2.78% during habituation; 2-3-1: the second and third members of a habituation triplet, followed by the first member of a different habituation triplet such that the TP and co-occurrence frequency between the second and third shapes of this test sequence were .33 and 2.78% during habituation; Figure 2-1C-D**).** Which test trial infants viewed first was randomized.

The only difference between the 3-1-2 and 2-3-1 part-sequences was whether the low-TP shape pair appeared in a sequence-initial or -final position. The only differences between the familiar and part-sequence test displays were the TPs between and co-occurrence frequencies of the shapes; this ensured that any looking time difference to the two types of test trials would necessarily be related to the statistical nature of the sequences. If infants segment the habituation sequence on the basis of statistical relations between shapes, they should discriminate between the familiar and part-sequence test displays. If infants preferentially track statistical regularities in sequence-initial or sequence-final positions, they should show differential discrimination between the familiar and part-sequence test displays in the 3-1-2 and 2-3-1 conditions.
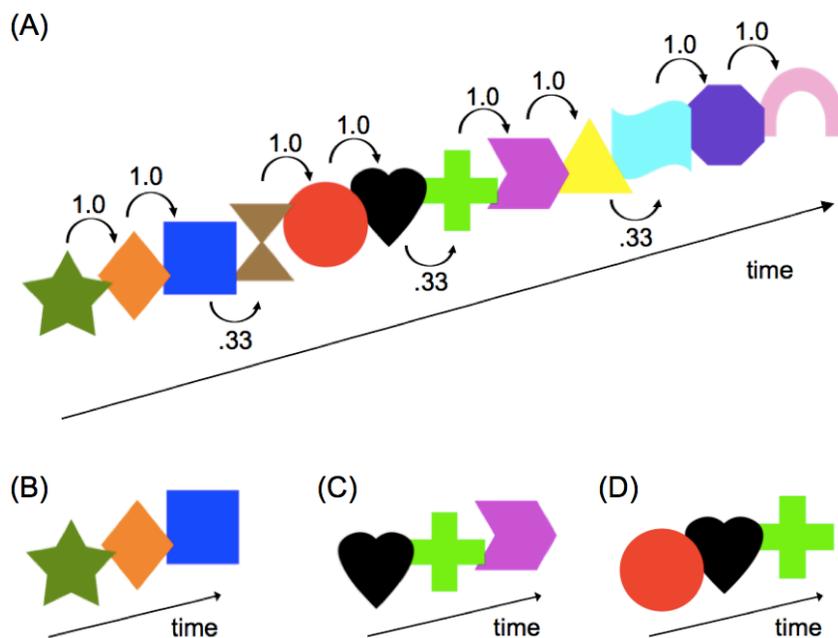
Figure 2-1. Habituation and test stimuli used in Experiment 1. Sample (A) habituation sequence, (B) familiar test sequence, (C) 3-1-2 test sequence, and (D) 2-3-1 test sequence. The heart is shown in black here but was white in the experiment.

Results

To examine infants' ability to discriminate the test sequences, we calculated each infant's novelty preference score: the proportion of total looking time on test trials that was directed to

53

the part-sequence test display. An independent-sample *t* test (this and all subsequent *t* tests were two-tailed) revealed that infants' mean novelty preference was not significantly different in the 3-1-2 (M = .508, SD = .092) and 2-3-1 (M = .485, SD = .123) conditions: $t(30)$ = 0.60, p = .551; therefore, we combined the data from both test conditions for subsequent analyses. As can be seen in Figure 2-2, infants' novelty preference was not significantly different from chance, where chance equals .50. Only 16 of the 32 infants showed a novelty preference score greater than .50 (one-sample Wilcoxon signed-rank test: Z = 0.21, p = .837), and infants' mean novelty preference (M = .496, SD = .108) was not significantly different from chance: $t(31)$ = 0.20, p = .842).
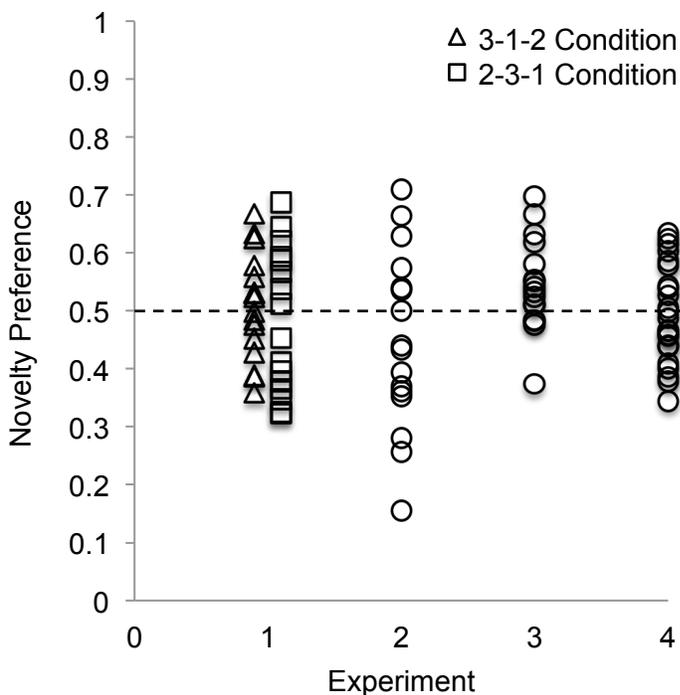
Figure 2-2. Individual infants' post-habituation novelty preference scores in Experiments 1-4. The habituation sequence consisted of either 12 shapes (Experiment 1) or 9 shapes (Experiments 2-4) organized in triplets. Test trials in Experiments 1, 2, and 4 contrasted a 1-2-3 triplet from habituation with a lower-TP part-triplet (both 2-3-1 and 3-1-2 part-triplets in Experiment 1; only 3-1-2 part-triplets in Experiments 2 and 4). Test trials in Experiment 3 contrasted the continuous, structured habituation sequence with a psuedo-randomly ordered sequence of the same shapes. Data are plotted separately for the 2-3-1 and 3-1-2 conditions of Experiment 1. Only Experiment 3 shows overall performance significantly different from chance.

Discussion

Experiment 1 was designed to examine whether infants' can acquire via VSL the same complexity of statistical structures that infants acquire via ASL. In contrast with previous findings of 8-month-olds' successful discrimination between familiar and part-sequences in ASL studies (e.g., Saffran et al., 1996a), the 8-month-olds in Experiment 1 did not show evidence of discriminating similar visual test sequences in both the 3-1-2 and 2-3-1 conditions. From these null results we are not able to draw conclusions about possible recency or primacy effects in infants' VSL, but infants' failure to show evidence of sensitivity to the statistical structure of the habituation sequence across both conditions suggests that the task presented a considerable challenge to infants, at least given the visual stimuli employed here.

It is unclear from Experiment 1 what was the cause of infants' failure to detect the statistical structure of the visual sequence. Studies demonstrating adults' facility at extracting TP information from visual sequences (e.g., Fiser & Aslin, 2002a) imply that learners must become capable of detecting statistical regularities in visual sequences at some point during development. Previous research suggests that one factor limiting infants' sensitivity to the statistical structure of visually-presented information is the number of items employed (e.g., Bulf et al., 2011; Téglás, Ibanez-Lillo, Costa, & Bonatti, 2015). This may be due to limitations in infants' visual attention and/or visual short-term memory. Thus, Experiment 2 was designed to test the possibility that reducing the complexity of the sequences, by decreasing the number of shapes used from 12 to 9, would enhance infants' VSL performance. Because infants performed similarly in the 3-1-2 and 2-3-1 conditions of Experiment 1, only a 3-1-2 condition was included in Experiment 2.

Experiment 2

Method

Participants. Sixteen healthy full-term 8-month-olds (7 females; *M* age 8 months 4 days, range = 7;17 to 8;15) were recruited as in Experiment 1. Data from an additional seven infants were excluded from the final sample due to fussiness (*n* = 4) or failure to habituate (*n* = 3).

Apparatus, stimuli, and procedure. The apparatus, stimuli, and procedure were identical to those of the 3-1-2 condition of Experiment 1, except only 9 (rather than 12) shapes were used. The habituation sequence therefore consisted of three pseudo-randomly ordered shape triplets (e.g. triplet 1: orange diamond, white heart, green plus; triplet 2: red circle, blue square, brown hourglass; triplet 3: yellow triangle, magenta arrow, olive star). As in Experiment 1, triplets could not repeat such that, in Experiment 2, TPs between shapes spanning triplet boundaries were .50 (rather than .33 as in Experiment 1). If infants are sensitive to the statistical structure of the habituation sequence, they should discriminate between the familiar and 3-1-2 test sequences.

Results and Discussion

Experiment 2 was designed to test the hypothesis that the complexity of the stimuli in Experiment 1 (i.e., the number of shapes) inhibited infants' VSL performance. However, similar to Experiment 1, infants' novelty preference in Experiment 2 was not significantly different from chance (see Figure 2-2). Only 7 of the 16 infants showed a novelty preference greater than .50 (one-sample Wilcoxon signed-rank test, $Z = 1.19$, $p = .234$), and infants' mean novelty preference (*M* = .450, *SD* = .155) was not significantly different from chance: $t(15) = 1.30$, $p = .215$. Thus, it does not appear that the reduction in the number of shapes from 12 to 9 had a significant effect on infants' VSL performances.

Together, the results of Experiments 1 and 2 suggest that detection of TP and/or co-occurrence frequency information in a visual shape stream is a difficult task for 8-month-olds. Yet previous research has asserted that by 8 months of age infants exhibit robust VSL ability, citing infants' ability to distinguish structured from random visual shape sequences (Kirkham et al., 2002). As mentioned previously, however, it is not clear from the Kirkham et al. method whether discrimination was a demonstration of SL, as performance was based on discriminating structured from random sequences – a task that infants can accomplish even without prior habituation to the shape stream (see Addyman & Mareschal, 2013).

Experiment 3 was designed to examine whether, with the stimuli employed in Experiment 2 in which infants show no evidence of sensitivity to TPs and/or co-occurrence frequencies, infants are still able to distinguish the structured habituation sequence from a random sequence. Positive evidence would suggest that infants may be sensitive to statistical differences between structured sequences and random sequences; however, this would also suggest that discrimination was not based on sensitivity to TPs and/or co-occurrence frequency information (based on the null results of Experiment 2), as it is in the auditory modality.

Experiment 3

Method

Participants. Sixteen healthy full-term 8-month-olds (7 females; *M* age 8 months 3 days, range = 7;19 to 8;22) were recruited as in Experiments 1 and 2. Data from an additional seven infants were excluded from the final sample due to fussiness (*n* = 3), parental interference (*n* = 2), sleepiness (*n* = 1), or failure to habituate (*n* = 1).

Apparatus, stimuli, and procedure. The apparatus, stimuli, and habituation procedure were identical to those of Experiment 2. The test procedure was modeled after Kirkham et al.

(2002). Infants viewed six test displays alternating between the familiar habituation shape stream and a novel stream produced by pseudo-randomly ordering the same nine habituation shapes; the single constraint on stimulus order in the random stream was that there were never two identical stimuli in a row. Thus, the familiar and random test streams could potentially be distinguished on the basis of a number of statistical differences (see Slone & Johnson, 2015). Both the familiar and random test streams were continuous, with no breaks or delays between shapes or triplets (Figure 2-3). Which test trial infants viewed first was randomized.

(A)

(B)

time

time
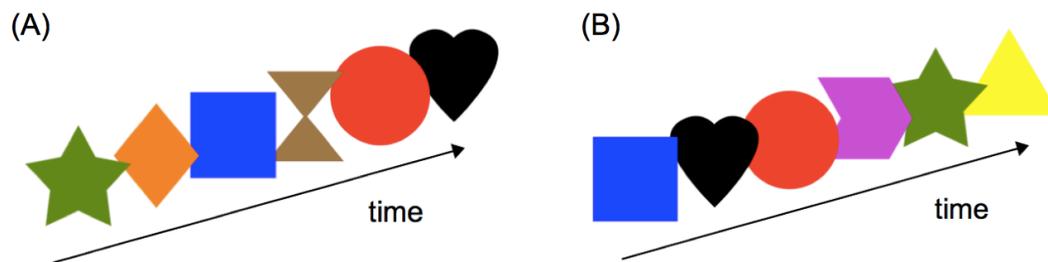
Figure 2-3. Test stimuli used in Experiment 3. The heart is shown in black here but was white in the experiment. (A) Sample familiar test sequence. (B) Sample random test sequence.

Results and Discussion

In contrast to Experiments 1 and 2, infants' novelty preference in Experiment 3 was significantly different from chance (see Figure 2-2). Twelve of the 16 infants showed a novelty preference greater than .50 (one-sample Wilcoxon signed-rank test, $Z = 2.22$, $p < .05$), and infants' mean novelty preference ($M = .546$, $SD = .080$) was significantly greater than chance: $t(15) = 2.29$, $p < .05$). Moreover, an independent-sample t test revealed that infants' mean novelty preference in Experiment 3 was significantly greater than infants' mean novelty preference in Experiment 2: $t(30) = 2.21$, $p < .05$.

Experiment 3 was designed to examine whether infants could discriminate the structured habituation sequence of Experiment 2 from a random sequence composed of the same nine

shapes. Although 8-month-olds showed no evidence of sensitivity to TPs or co-occurrence

frequencies after habituation in Experiment 2, the 8-month-olds in Experiment 3 were able to

distinguish the habituation sequence from a random sequence. These findings suggest that

performance on the discrimination task in Experiment 3 did not depend on sensitivity to TPs or

co-occurrence frequencies.

Together, the present series of studies suggests that detecting and responding to

differences in TPs and co-occurrence frequencies in triplet-based visual sequences is particularly

challenging for 8-month-olds. It is unclear why this is the case, especially given multiple

demonstrations of infants' sensitivity to TPs in auditory tasks employing very similar methods.

One possibility is that the testing methods used in ASL tasks (familiarization and head-turn) are

incommensurate with the testing methods used in VSL tasks (visual habituation). However,

visual habituation has been shown to be an effective measure of learning in thousands of studies

with infants. Therefore, in Experiment 4, we investigate the alternate possibility that, unlike what

has been previously suggested, VSL is limited compared to ASL in infancy. If this is the case, it

is important to understand the factors that may be constraining infants' VSL performance.

One hypothesis proposed to explain infants' difficulty with VSL is that constraints in

cognitive processing ability, particularly visual short-term memory (VSTM), may be a limiting

factor (Bulf et al., 2011). Although no published studies have yet empirically investigated this

possibility, the hypothesis makes intuitive sense: It is rare to find a real-world situation in which

the visual system is required to process rapidly presented sequential information; however, the

auditory system processes such information daily (e.g., speech). Thus, unlike the auditory

system, the visual system is not designed to encode and retrieve rapid sequential information,

possibly constraining visual sequence learning (Conway & Christiansen, 2009; Emberson et al., 2011).

Experiment 4 was designed to investigate this possibility by examining the relation between infants' VSTM and VSL performances. In addition, infants in Experiment 4 also participated in a general developmental assessment, the Mullen Scales of Early Learning (MSEL; Mullen, 1995) to probe the relation between VSL and cognitive development more generally. Including the MSEL also allowed us to examine whether potential relations between VSL and VSTM performances remained significant after controlling for infants' overall cognitive ability.

Experiment 4

Method

Infants participated in three tasks, an infant-controlled habituation task identical to Experiment 2 to test infants' VSL, a 'one-shot' VSTM task designed by Oakes, Baumgartner, Barrett, Messenger, and Luck (2013), and the MSEL (Mullen, 1995).

Participants. Twenty-three healthy full-term 8-month-olds (8 females; *M* age 7 months 29 days, range = 7;0 to 8;18) were recruited as in Experiments 1-3. Data from an additional 17 infants were excluded from the final sample due to failure to habituate (*n* = 8), fussiness (*n* = 6), experimenter error (*n* = 2), or premature birth (*n* = 1).

Apparatus and stimuli. The apparatus and stimuli for the VSL task were identical to those of Experiment 2. For the VSTM task, an Eyelink 1000 eye tracker (SR Research) with a 55.9-cm color monitor displayed stimuli and collected eye-tracking data. A PC computer running Experiment Builder software controlled stimulus presentation and sent markers stored with eye tracker data, allowing us to coordinate infants' eye movements with the stimuli. The eye-tracking

system recorded point-of-gaze (POG) coordinates (spatial resolution within 1.0° visual angle) at 500 Hz.

Each VSTM trial consisted of a sample array of two 9° x 9° colored squares centered 8° from the center of the monitor and presented on a gray background. The two colors were selected at random from a set of 8 colors: black, blue, brown, cyan, green, gold, pink, and purple. These colors were different from those of the VSL task to reduce the possibility of carry-over effects. After 500 ms, the two squares disappeared for 300 ms. Two squares then reappeared in the same locations, but the color of one square had changed to one of the remaining six colors. This test array remained visible for 1500 ms until the trial ended (Figure 2-4). The experimental script randomly selected the colors of the two squares in the sample array, determined whether the changing stimulus appeared on the left or right, and randomly selected the color for the changed square from the remaining six colors; the only constraint was that the changing stimulus could not be on the same side of the screen more than three times in a row. Each infant received a different order of stimuli, randomly determined within the constraints of the experimental design. Each trial lasted approximately 2300 ms.

Procedure. All infants participated in the VSL task first. The procedure for the VSL task was identical to that of Experiment 2. Infants then took a 5-mintue break while the experimenter set up for the VSTM task. Interpreting VSL performance depended on completion of the entire task (as the critical "test" trials came at the very end of the task), whereas it was not critical (or expected) that infants would complete all 90 VSTM trials. Thus, infants completed the VSL task first to ensure that, if an infant became fussy and could not complete the VSTM task, we still had usable data from both tasks.
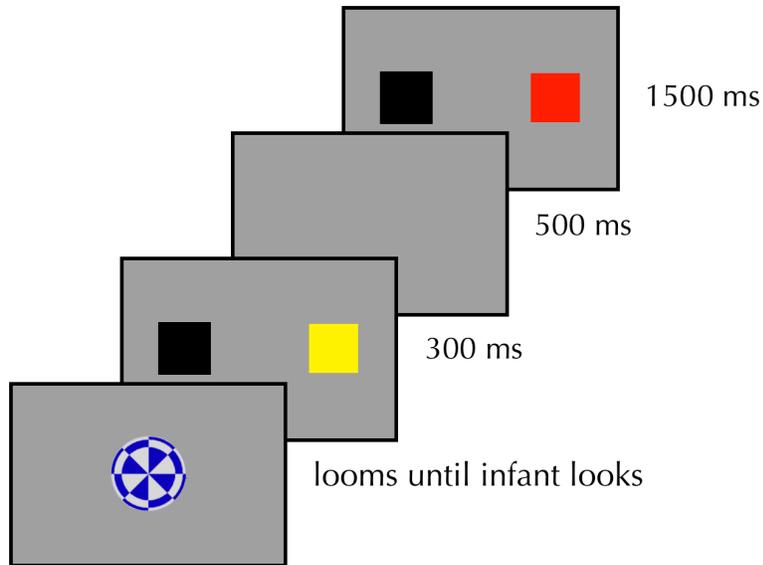
Figure 2-4. Sample sequence of events in the visual short-term memory task.

During the VSTM task, infants again sat on a caregivers' lap approximately 60 cm from the eye-tracking monitor. POG was calibrated using Experiment Builder software. A series of small expanding and contracting circular stimuli accompanied by sounds were presented to attract infants' attention to five locations that formed a crosshairs shape on the screen (bottom center, center, top center, center left, and center right). After the best possible calibration was obtained, caretakers were again instructed not to interact with their infant or attend to the monitor. An attention-getter was then presented to attract infants' attention to the center of the screen. A trained observer not visible to the infant viewed the infant's face on a television, fed by a signal from a video camera situated below the monitor presenting the stimulus. Each VSTM trial began when the experimenter pressed a key indicating that the infant was looking at the attention getter. The experimental session continued until the infant became fussy or stopped looking at the monitor, or until 90 VSTM trials had elapsed. Infants completed an average of 70 VSTM trials ($SD = 20$) and provided usable eye-tracking data on an average of 62 trials ($SD = 20$).

All infants then completed the MSEL last because pilot data suggested that this was the task during which infants were least likely to become fussy. The MSEL was administered in the standardized format, with infants sitting in a caregiver's lap at a table for the VR and FM tasks, and sitting on the floor for the RL tasks. The number of infants contributing useable data on each of the VSL, VSTM, and MSEL tasks is listed in Table 2-1.

Coding. For the VSTM task, two Areas of Interest (AOIs) were defined that corresponded to the test array's changed and unchanged colored squares. AOIs were centered on the locations of the two squares and were 11° by 11° to allow for 1° of imprecision in the eye-tracker's localization of infants' POG. The number and duration of fixations that fell within these AOIs were calculated for each infant using SR Research's velocity-based algorithms. If, during the test array, an infant fixated the changed color reliably more than the unchanged color, we had strong evidence that they formed a VSTM representation of at least one item from the sample array. To calculate each infant's preference for the changed color, we divided duration of looking within the AOI for the changed item by the duration of looking within both AOIs combined, for each trial. This change preference score was .50 (chance) if an infant looked equally to the changed and unchanged items, and was >.50 if an infant preferred the changed item. We calculated mean change preference score across trials both for individual infants and for infants as a group.

From the MSEL we obtained data on five subscales: Expressive Language (EL), Receptive Language (RL), Visual Reception (VR), Fine Motor (FM), and Gross Motor (GM). From these measures, standard scores for overall cognitive ability, non-verbal cognitive ability (NVCOG), and verbal cognitive ability (VCOG) were calculated for each infant and used for analysis. Based on prior literature, VCOG was calculated using the average of the EL and RL

63

Table 2-1

Descriptive Statistics for the VSL, VSTM, and MSEL tasks in Experiment 4

Task: Dependent Variable	N	Mean (SD)	Range
VSL: Novelty Preference	23	.490 (.094)	.285-.635
VSTM: Change Preference	22	.555 (.060)	.416-.651
MSEL: Receptive Language (RL)	16	53.1 (8.4)	38-62
MSEL: Expressive Language (EL)	16	59.4 (10.2)	31-69
MSEL: Visual Reception (VR)	16	53.4 (5.4)	44-61
MSEL: Fine Motor (FM)	16	52.9 (8.8)	38-68
MSEL: Nonverbal Cognitive Ability (NVCOG)	16	53.2 (5.0)	41.0-64.5
MSEL: Verbal Cognitive Ability (VCOG)	16	56.3 (5.3)	46.5-63.5
MSEL: Overall Cognitive Ability	16	54.7 (3.3)	49.75-61.25

subscale T scores, and NVCOG was calculated using the average of the VR and FM subscale T scores (Jeste et al., 2014). Overall cognitive ability was calculated by averaging the T scores of the EL, RL, VR, and FM subscales.

Results

VSL. Twenty-three 8-month-olds completed the VSL task. Similar to Experiment 2, infants' novelty preference in Experiment 4 was not significantly different from chance (see Figure 2-2). Only 10 of the 23 infants showed a novelty preference greater than .50 (one-sample Wilcoxon signed-rank test, $Z = 0.41$, $p = .681$), and infants' mean novelty preference (see Table 2-1) was not significantly different from chance: $t(22) = 0.50$, $p = .621$. Thus, as a group, infants did not show evidence of successful VSL.

VSTM. Twenty-three 8-month-olds completed the VSTM task; however, to reduce the likelihood of including colorblind infants in our sample, we excluded from the analyses reported here one infant whose maternal familial colorblindness history put him at significant risk of colorblindness.

Infants looked significantly longer to the changed color relative to the unchanged color during test arrays. Eighteen of the 22 infants had a change preference score >.50, which was significantly different from chance (one-sample Wilcoxon signed-rank test, $Z = 3.20$, $p = .001$). Moreover, as a group, infants' mean change preference was significantly different from chance, $t(21) = 4.25$, $p < .001$, where chance equals 0.50 (see Table 2-1). These results suggest that, both individually and as a group, infants formed VSTM representations of at least one item from the sample arrays and maintained these representations across the 300 ms delays between the sample and test arrays.

MSEL. Sixteen 8-month-olds completed the MSEL. Averages and standard deviations for MSEL subscales and standard scores are shown in Table 2-1.

Relations between measures. Experiment 4 was designed to investigate possible factors constraining VSL performance. Although infants did not show evidence of VSL as a group, there was considerable variability in infants' novelty preference scores (see Table 2-1). Thus, we examined whether individual differences in VSL performance were related to individual differences in VSTM and MSEL measures. Significant positive correlations between VSL and VSTM and/or MSEL measures would suggest that these factors may be constraining VSL performance at this age.

VSL and VSTM. We first examined the correlation between infants' VSL and VSTM performances. We hypothesized that VSTM may be a critical factor supporting infants' VSL,

such that we would find a significant positive relation between VSL and VSTM. Contrary to our predictions, however, VSL and VSTM were not significantly correlated: $r = -.243$, $p = .276$.

  VSL and MSEL. Next, we examined the relation between VSL and infants' overall cognitive ability, NVCOG, and VCOG, as measured by the MSEL. We hypothesized that cognitive ability might relate to and possibly inform differential VSL performance. Given the data from Experiments 1-3 suggesting that infants may not acquire via VSL the triplet structures that can be acquired via ASL, we were interested in whether VSL performance would be differentially related to VCOG versus NVCOG. That is, if differential VSL and ASL performances in infancy are due to modality constraints on SL, we hypothesized that we may find differential relations between VSL and VCOG versus NVCOG (Jeste et al., 2014). If, in contrast, SL is not modality specific, we hypothesized that we may find significant relations between VSL and overall cognitive ability, and similar relations between VSL and VCOG versus NVCOG. When analyzing the relations between VSL performances and MSEL measures, we controlled for infants' age in months. This was done to control for the facts that (1) the ages of the participants in Experiment 4 ranged from 7 months 0 days to 8 months 18 days, and (2) MSEL standard scores are computed using different norms for 7-month-old (operationalized as 7 months 0 days to 7 months 30 days) versus 8-month-old (operationalized as 8 months 0 days to 8 months 30 days) infants (Mullen, 1995).

  As can be seen in Figure 2-5, VSL novelty preference was positively correlated with all three VCOG, NVCOG, and overall cognitive ability. Only the relation between novelty preference and overall cognitive ability was statistically significant, controlling for age: novelty preference and cognitive ability, $r = .535$, $p < .05$; novelty preference and VCOG, $r = .338$, $p = .237$; novelty preference and NVCOG, $r = .334$, $p = .243$. The relations between VSL and these
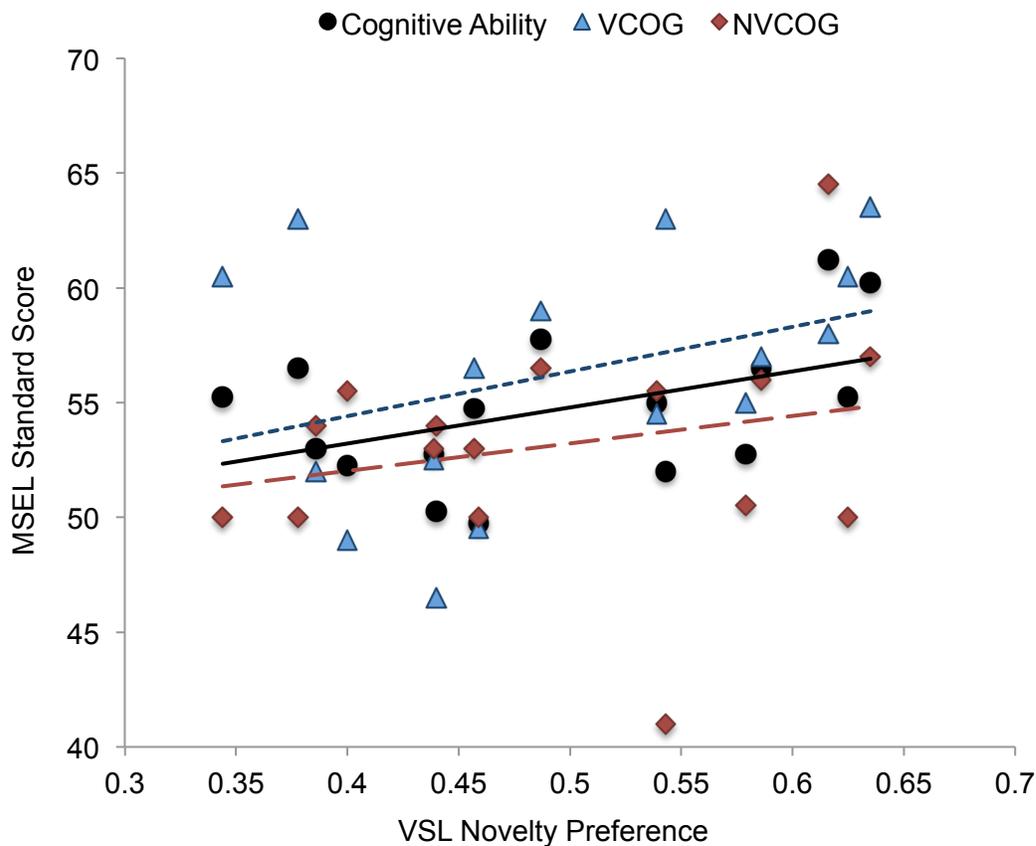
Figure 2-5. Individual infants' overall cognitive ability, verbal cognitive ability (VCOG), and nonverbal cognitive ability (NVCOG) MSEL scores, plotted against infants' VSL novelty preference in Experiment 4.

three MSEL measures were not significantly different from one another: Hotelling's ts(13) < 1.04, ps > .319.

VSL, VSTM, and MSEL. A secondary purpose for including the MSEL in Experiment 4 was to examine potential relations between VSL and VSTM, controlling for infants' overall cognitive ability. However, when controlling only for overall cognitive ability, or for both overall cognitive ability and age, the relation between VSL and VSTM remained non-significant: rs < -.233, ps > .424.

Discussion

Experiment 4 was designed to investigate possible factors constraining VSL performance in infancy by having infants participate in a VSTM task and the MSEL, in addition to the VSL task from Experiment 2. Similar to Experiment 2, after habituating to a continuous sequence of shapes organized into triplets, infants did not discriminate between high- and low-probability shape sequences, suggesting that infants as a group were not sensitive to the statistical structure of the habituation sequence. We also examined whether individual differences in novelty preference were related to individual differences in VSTM and MSEL measures. Specifically, we hypothesized that VSTM may be a critical factor supporting infants' VSL, such that we would find a significant positive relation between VSL and VSTM. Infants showed evidence of successful VSTM for colored shapes, both individually and as a group, however contrary to our predictions, VSTM and VSL performances were not significantly correlated.

The lack of a significant relation between measures of VSL and VSTM suggests that VSTM may not be a primary factor constraining VSL in infancy. Though some degree of VSTM must be necessary for VSL ability, it could be that by 8 months of age infants have sufficient VSTM ability such that VSTM performance is no longer the primary factor limiting VSL performance. That the 8-month-olds in Experiment 4 exhibited strong evidence of VSTM both individually and as a group seems to support this possibility.

It is also possible, however, that VSTM may be an important factor constraining infants' VSL performance, but that the VSTM task employed in Experiment 4 did not allow us to detect this relation. For instance, it is possible that we did not detect a relation between VSL and VSTM performances simply due to infants' overall strong VSTM performance and the accompanying dearth of individual differences in VSTM compared to VSL performances (see Table 2-1).

Alternatively, the VSTM task may have been a weak proxy for its intended construct. We used the Oakes et al. (2013) 'one-shot' VSTM task because the memory demands the task imposes on infants appear highly similar to those imposed by the VSL task – namely, representing colored shapes in memory for brief periods of time and constantly encoding new memories as new shapes are presented. However, there may be important memory demands in the VSL task that were not required by the VSTM task, such as encoding the shapes in addition to the colors of items, binding representations of items presented over time together in memory, or retaining the specific sequences of items. Future research could examine whether, with a more difficult or targeted VSTM task that may yield more meaningful individual differences in performance, a relation is found between VSTM and VSL performances.

In addition to examining the role of VSTM, we also hypothesized that cognitive ability might relate to and possibly inform differential VSL performance. Specifically, we hypothesized that VSL performance might either (1) be differentially related to VCOG versus NVCOG, supporting the domain specificity of SL ability in infancy, or (2) be related to overall cognitive ability and show similar relations to both VCOG and NVCOG, supporting the domain generality of VSL.

Our findings aligned with our second hypothesis: VSL novelty preference was positively correlated with all three VCOG, NVCOG, and overall cognitive ability, though only the correlation with overall cognitive ability was significant (controlling for age). These findings accord with previous research showing associations between visual perceptual learning ability and both concurrent and later cognitive skills (e.g., Colombo, Richman, Shaddy, Maikranz, & Blaga, 2004; Rose, Feldman, & Jankowski, 2009; Shafo, Conway, Field, & Houston, 2012; Thompson, Fagan, & Fulker, 1991). That VSL was not differentially related to VCOG versus

NVCOG suggests that the processes underlying VSL ability may be similar to those underlying both verbal and nonverbal cognitive ability, suggesting that VSL performance may reflect domain general learning processes in early cognitive development. These results are limited by the correlational nature of Experiment 4, and by the relatively low number of participants that completed all three tasks. Nonetheless, the relations among measures reported here suggest that individual differences in VSL performance may be meaningful in terms of both VSL ability in particular and early cognitive development more broadly.

General Discussion

The present series of experiments was designed to address two related questions. First, we asked whether infants acquire via VSL the same complexity of statistical structures that infants acquire via ASL. Specifically, we employed a triplet sequence with which 8-month-olds successfully detect statistical relations among aurally presented items, and asked: Do infants detect similar statistical relations when the triplet structure is instantiated in a visual shape sequence? Second, we examined possible factors constraining VSL in infancy. Specifically, we asked: Does infants' VSL performance correlate with concurrent visual short-term memory and/or general cognitive ability?

Altogether, the present series of experiments suggests that, although 8-month-olds are able to distinguish triplet-structured from random visual sequences, detecting and responding to differences in TPs in triplet-based visual sequences is particularly challenging. The difference in infants' performance on the random versus structured test (Experiment 3) compared to the high-versus low-TP tests (Experiments 1, 2, and 4) is important for understanding the extent and utility of VSL in infancy. Although discriminating structured from random sequences may involve sensitivity to some statistical information in the sequences, this discrimination does not

70

appear to result from sensitivity to conditional relations between items. Conditional relations are particularly informative statistics because they reflect the predictive power of one event with respect to another, and therefore may be critical not only for associative learning, but also for inference (Fiser & Aslin, 2002a; Krogh, Vlach, & Johnson, 2013). Insensitivity to conditional relations between visual items at an age at which infants show sensitivity to these relations between auditory items suggests that the information that can be learned via VSL may be limited compared to what can be learned via ASL.

Experiment 4 examined the possibility that VSTM ability may constrain VSL at 8 months of age. Infants exhibited strong evidence of VSTM for colored shapes, both individually and as a group, however no relation was found between VSTM and VSL performances, suggesting that VSTM is not the primary factor constraining 8-month-olds' VSL ability.

An alternate possibility is that VSL and ASL abilities may be subserved by different mechanisms in infancy. This possibility is consistent with domain- and modality-specific views of SL (e.g., Conway & Christiansen, 2006). We examined possible differences in the relation between VSL and verbal versus non-verbal cognitive abilities in Experiment 4, to examine the domain generality versus specificity of VSL. VSL was positively correlated with both verbal and nonverbal cognitive abilities, suggesting that the processes underlying VSL are common to both of these domains. Nevertheless, at 8 months of age, the verbal cognitive abilities measured by the MSEL are arguably proto-linguistic abilities. It is possible that, if verbal cognition were tested at an older age at which measures of verbal cognition were more dependent on infants' language production, the association between VSL and verbal cognition may differ from what was observed in the present study. Recent findings of relations between visual implicit learning and language ability in both adults (Conway, Bauernschmidt, Huang, & Pisoni, 2010) and infants

71

(Ellis, Gonzalez, & Deák 2014; Shafto, Conway, Field, & Houston, 2012), however, suggest that the present findings are not simply a reflection of the MSEL verbal subscales assessing proto-linguistic abilities at 8 months of age.

Other possible explanations of infants' limited VSL performance compared to what has been demonstrated previously with ASL tasks relate to the stimuli employed in these different tasks. The stimuli used in the present VSL task differ from typical ASL stimuli in at least three ways that could affect infants' SL performances: (1) accord with the dominant presentation format outside the laboratory, (2) presentation rate, and (3) familiarity to infants. First, research suggests that SL operates optimally over the types of input that occur most frequently in the environment (see Krogh et al., 2013 for a review). In learners' everyday environments, auditory information is conveyed temporally whereas visual information is primarily conveyed spatially. Research by Conway and Christiansen (2009) suggests that adults learned an artificial grammar presented in the auditory or visual modalities equally well when auditory input was presented temporally and visual input was presented spatially. Adults learned the grammar significantly less well when visual input was presented temporally. Thus, it is possible that infants performed poorly in the present VSL task because shapes were presented sequentially rather than spatially. Indeed, data from Kirkham et al. (2007) suggests that adding spatial information to a sequential VSL task may improve infants' performance. The present series of experiments employed sequentially presented visual stimuli because (1) this is the presentation format employed by the majority of previous infant VSL studies, (2) we designed our task to be as similar as possible to the Saffran et al. (1996a) ASL task, and (3) the present study was designed to examine the limits of infants' VSL ability. An intriguing question for future research is whether infants show successful VSL of triplet structures when stimuli are arrayed spatially.

A second way in which our VSL stimuli differed from typical ASL stimuli is presentation rate. Auditory items in the Saffran et al. (1996a) study were presented at a rate of one syllable every 222 ms. In contrast, visual items in the present study were presented at a rate of one shape every 750 ms. Shapes were presented at this relatively slow rate because of the looming nature of the stimuli, which helped maintain infants' visual engagement with the shape stream. However, as a consequence of this slower presentation rate, infants saw fewer visual stimuli and therefore received less statistical evidence during habituation in the present series of experiments (an average of 13 tokens of each triplet in Experiment 1, and 17 tokens of each triplet in experiments 2 and 4) compared to familiarization in the Saffran, et al. study (45 tokens of each triplet). Thus, it is possible that infants in the present series of VSL experiments simply lacked sufficient statistical information to allow successful VSL performance.

A third way in which our VSL stimuli differed from typical ASL stimuli is in terms of familiarity to infants. ASL studies typically employ speech sounds, whereas the present VSL experiments employed geometric shapes. By 8 months of age, infants have likely had considerably more exposure to speech sounds than geometric shapes, making VSL stimuli less familiar than the typical ASL stimuli. Interestingly, recent studies report 7- to 9-month-olds' sensitivity to TPs in visual event sequences involving human actors Roseberry et al. (2011) and cartoon-like characters (Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014). These studies suggest that infants may indeed be capable of acquiring relatively complex statistical structures via VSL when stimuli are more familiar and possibly more engaging than those employed in the present series of experiments. It remains unclear whether infants' performances in the Stahl et al. (2014) and Roseberry et al. (2011) studies were facilitated by the familiarity of the stimuli, by the complexity and attention-engaging nature of the stimuli, or both. Ongoing

work in our lab is currently examining the contributions of these two factors to VSL performance, to better understand the factors that may facilitate VSL in infancy.

In sum, the present series of studies were designed to be as similar as possible to the Saffran et al. (1996a) ASL experiments, to examine whether infants can acquire via VSL structures similar to those acquired via ASL. In contrast to previous findings in the auditory modality, we found no evidence of infants' sensitivity to TP statistics in visual triplet structures. It is possible that, had we employed visual stimuli other than sequences of shapes, we may have found evidence of successful VSL; however, the purpose of the present series of experiments was to examine the limits of and possible factors constraining VSL in infancy. Our results have important theoretical implications for SL ability across modalities in infancy and the mechanisms that may underlie SL performance. Future research should continue to examine both the limits of and the processes that facilitate SL in infancy.

References

Addyman, C., and Mareschal, D. (2013). Local redundancy governs infants' spontaneous orienting to visual‑temporal sequences. *Child development*, *84*, 1137-1144. DOI: 10.1111/cdev.12060

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9, 321-324. DOI:10.1111/1467-9280.00063

Bates, E., & Elman, J. (1996). Learning rediscovered: A perspective on Saffran, Aslin, and Newport. *Science*, *274*, 1849-1850.

Bulf, H., Johnson, S. P., and Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121,* 127-132. DOI:10.1016/j.cognition.2011.06.010

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., and Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy*, *5*, 1-38. DOI:10.1207/s15327078in0501_1

Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356-371. DOI: 10.1016/j.cognition.2009.10.009

Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24-39. DOI:10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905-912. DOI:10.1111/j.1467-9280.2006.01801.x

Conway, C. M. and Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology* 21, 561-580. DOI:10.1080/09541440802097951

Ellis, E. M., Gonzalez, M. R., and Deák, G. O. (2014). Visual prediction in infancy: what is the association with later vocabulary?. *Language Learning and Development, 10*, 36-50. DOI: 10.1080/15475441.2013.799988

Emberson, L. L., Conway, C. M., and Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology* 64, 1021-1040. DOI:10.1080/17470218.2010.538972

Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science* 12, 499-504. DOI:10.1111/1467-9280.00392

Fiser, J., and Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning Memory and Cognition* 28, 458-467. DOI:10.1037/0278-7393.28.3.458

Fiser, J. & Aslin, R.N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99,* 15822-15826. DOI:10.1073/pnas.232472899

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254-260. DOI:10.1111/j.1467-9280.2007.01885.x

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., ... and Johnson, S.

P. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science*, *18*, 90-105. DOI: 10.1111/desc.12188

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42. DOI:10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., and Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78,* 1559-1571. DOI:10.1111/j.1467-8624.2007.01083.x

Krogh, L., Vlach, H. A., and Johnson, S. P. (2012). Statistical learning across development: flexible yet constrained. *Frontiers in psychology*, *3,* 598. DOI: 10.3389/fpsyg.2012.00598

Marcovitch, S. and Lewkowicz, D. L. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science, 12,* 1020-1025. DOI:10.1111/j.1467-7687.2009.00838.x

Mullen, E. M. (1995). *Mullen scales of early learning* (pp. 58-64). Circle Pines, MN: AGS.

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology, 4*, 697. DOI: 10.3389/fpsyg.2013.00697

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80,* 674-685. DOI:10.1111/j.1467-8624.2009.01290.x

Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: Getting some help from the auditory modality. In *Proceedings of the 29th Annual Cognitive Science Society* (pp. 611-616). Austin, TX: Cognitive Science Society.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development*, *80*, 134-150. DOI: 10.1111/j.1467-8624.2008.01250.x

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R., and Shipley, T. (2011). Babies catch a break: 7- to 9-month-olds track statistical probabilities in continuous dynamic events. *Psychological Science, 22,* 1422-1424. DOI:10.1177/0956797611422074

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*, 149-169. DOI: 10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172-196. DOI:10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928. DOI:10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27-52. DOI:10.1016/S0010-0277(98)0075-4

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. DOI: 10.1016/S0010-0277(98)00075-4

Saffran, J. R. and Wilson, D. P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy, 4,* 273-284. DOI:10.1207/S15327078IN0402_07

Shafto, C. L., Conway, C. M., Field, S. L., and Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*, 247-271. DOI:10.1111/j.1532-7078.2011.00085.x

Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2-and 5-month-olds'

    segmentation of continuous visual sequences. *Journal of experimental child psychology*,

    133, 47-56. DOI:10.1016/j.jecp.2015.01.007

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., and Hirsh-Pasek, K. (2014).

    Infants segment continuous events using transitional probabilities. *Child Development, 85,*

    1821-1826. DOI: 10.1111/cdev.12247

Téglás, E., Ibanez-Lillo, A., Costa, A. and Bonatti, L. L. (2015), Numerical representations and

    intuitions of probabilities at 12 months. *Developmental Science, 18*, 183-193.

    DOI: 10.1111/desc.12196

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word

    segmentation. *Infancy, 7,* 53-71. DOI:10.1207/s15327078in0701_5

Thompson, L. A., Fagan, J. F., and Fulker, D. W. (1991). Longitudinal prediction of specific

    cognitive abilities from infant novelty preference. *Child Development*, *62*, 530-538.

    DOI: 10.1111/j.1467-8624.1991.tb01549.x

PAPER 3:

Statistical and chunking processes in adults' visual statistical learning

Abstract

Much research has documented learners' ability to segment auditory and visual input into its component units. Two types of models have been designed to account for this phenomena: statistical models, in which learners represent statistical relations between elements, and chunking models, in which learners represent statistically coherent units of information. In a series of three experiments, we investigated how adults' performance on a visual sequence-learning task aligned with the predictions of these two types of models. Experiments 1 and 2 examined learning of embedded items and Experiment 3 examined learning of illusory items. The pattern of results obtained was most consistent with the competitive chunking model of Servan-Schreiber and Anderson (1990). Implications for theories and models of statistical learning are discussed.

Statistical and Chunking Processes in Adults' Visual Statistical Learning

The means by which humans acquire and represent knowledge is fundamental to cognitive science. One important mechanism shown to support learning across domains is learners' ability to detect statistical associations among elements in a sensory array (e.g., Fiser & Aslin, 2001, 2002; Saffran, Newport, & Aslin, 1996a). Notably, this statistical learning (SL) ability has been demonstrated across the lifespan (e.g., Saffran, Aslin, & Newport, 1996b) and even across species (e.g., Toro & Trobalón, 2005).

Despite the scope of SL, the processes underlying SL remain unclear. Traditionally, SL has been conceptualized as sensitivity to statistical relations among elements. For instance, in their seminal studies of statistical word segmentation, Saffran et al. (1996a,b) exposed participants to a continuous stream of speech in an artificial language. After very limited exposure, participants showed evidence of segmenting the stream into its component words. Saffran et al. proposed that this ability might have resulted from computation of transitional probabilities (TPs) between syllables in the stream. Transitional probability (TP) is defined as the probability of event Y given event X, and is a measure of the strength with which X predicts Y. Saffran et al. hypothesized that learners could track TPs between adjacent syllables, using peaks in TP to group syllables into words, and dips in TP to identify breaks between words.

The view that SL occurs via computations has prevailed in literatures on auditory artificial language learning (e.g. Saffran et al., 1996a,b) and visual sequence (e.g., Fiser & Aslin, 2002) and scene learning (e.g., Fiser & Aslin, 2001). Models instantiating such computational approaches to segmentation are typically SRNs (e.g., Elman, 1990), which learn and represent statistical relations between elements, but do *not* represent the units they segment.

Recently, there have been attempts to account for word segmentation with a different

type of model: chunking models (e.g., Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Orbán, Fiser, Aslin, & Lengyel, 2008; Perruchet & Vinter, 1998), which propose that learners *do* represent the statistically coherent "chunks" of information from the input. Perruchet and Vinter's (1998) PARSER model, for instance, assumes that elements perceived within one attentional focus are "chunked" into a new, larger representation. Representations of chunks presented repeatedly are strengthened in memory; chunks presented rarely are forgotten. Applied to Saffran et al.'s (1996a,b) task, PARSER claims that representations of chunks within words are strengthened (because they are repeated more frequently), while chunks spanning word boundaries are forgotten. Thus, chunking models like PARSER predict that segmentation operates according to very different means (representing units) than those proposed by statistical models (representing statistical relations, *not* units).

It is unclear which type of model best accords with how learners process and represent information. Recent studies have been designed to distinguish between these models by examining situations in which chunking and statistical models make contrasting predictions. These studies examine learning of (1) embedded items and (2) illusory items.

Embedded items are features that occur only within larger features (Fiser & Aslin, 2005). Statistical models assume learners represent statistical relations between all pairs of adjacent elements such that, as learners become familiar with a unit, distinguishing components embedded in that unit improves relative to random configurations of elements (see Giroux & Rey, 2009). Many chunking models, in contrast, assume that as learners become familiar with a unit, they become *less* able to distinguish components embedded in that unit from random configurations of elements (see Giroux & Rey, 2009). That is, representations of embedded items and their larger units compete in memory. Over time, memory for the unit gets strengthened

while competing representations of embedded items vanish.

These contrasting predictions concerning embedded items have been empirically tested with adults in auditory artificial language learning tasks (Giroux & Rey, 2009) and in visual scene learning tasks (Fiser & Aslin, 2005). Results from both studies align with the predictions of chunking models: while learners distinguish units (e.g., words) from random configurations of sounds or shapes, they are unable to distinguish embedded units from random configurations.

Illusory items are items that are never presented to participants, but have the same statistical structure as other items that are presented. For example, if *tazepi*, *mizeru*, and *tanoru,* are words presented in a speech stream, and TPs are .50 between syllables within these words (e.g., between *ta* and *ze* and between *ze* and *ru*), a statistically matched illusory word would be *tazeru* (Endress & Mehler, 2009). If learners only represent statistical relations between elements, words and illusory words should be indistinguishable. If learners chunk and represent entire words, however, they should fail to recognize illusory words. In an auditory artificial language learning task Endress and Mehler (2009) found that, while participants distinguished words from lower-TP "part-sequences," they did not distinguish words from illusory words, suggesting that they represented statistical relations, rather than chunks.

Thus, studies of embedded and illusory items have yielded conflicting evidence regarding whether learners represent statistical relations or chunks. However, these studies employed widely varying methods, making it difficult to determine whether differences in performance across tasks were due to different underlying processes, or simply to methodological differences between studies.

The goal of the present series of experiments was to overcome this limitation and to extend previous work by investigating learning of both embedded (Experiments 1 and 2) and

illusory (Experiment 3) items in a visual sequence-learning (VSL) task. There are three main contributions of this work: (1) we contribute a variety of new human data about VSL under a range of experimental conditions; (2) we examine learning of both embedded and illusory items using highly comparable methods across tasks; and (3) we consider how the data fit with a variety of statistical and chunking models.

Experiment 1

Method

Participants. Thirty-six undergraduate students were recruited from Psychology classes at the University of California, Los Angeles. Participants were randomly assigned to participate in either a 10-minute ($N = 18$; 15 females; M age 20.2 years; range = 18.6 to 21.9) or 20-minute ($N = 18$; 14 females; M age 20.9 years; range = 18.5 to 29.1) familiarization condition. Data from an additional 16 participants were excluded from the final sample due to poor calibration or insufficient eye tracking data ($n = 10$), eye tracker failure ($n = 1$), or sleepiness ($n = 5$). All participants earned course credit for their participation.

Apparatus and stimuli. An Eyelink 1000 eye tracker with a 55.9-cm color monitor displayed stimuli and collected eye-tracking data. A PC computer running Experiment Builder software controlled stimulus presentation and sent markers stored with eye tracker data, allowing us to coordinate participants' eye movements with the stimuli. The eye-tracking system recorded point-of-gaze (POG) coordinates (spatial resolution within 1.0° visual angle) at 500 Hz.

Stimuli were 10 colored shapes on a black background (Figure 3-1). Each shape loomed for 750 ms within one of 10 grid locations on the monitor. Shapes were randomly organized for each participant into four units: two triplets and two pairs (e.g. triplet 1: star, diamond, square; triplet 2: hourglass, circle, heart; pair 1: plus, arrow; pair 2: triangle, banner). For simplicity, the
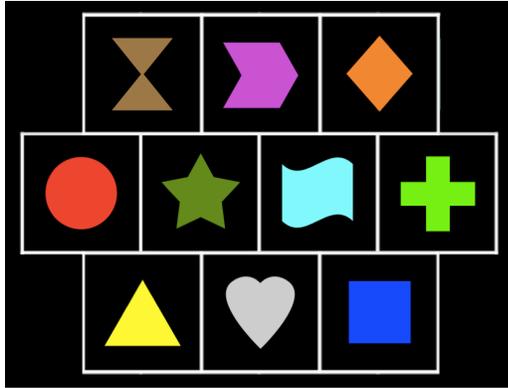
Figure 3-1. Sample stimulus array used in Experiment 1. Only one shape appeared on the screen at a time.

10 shapes will be referred to by the letters ABCDEFGHIJ, where 'ABC' and 'DEF' are the two triplets and 'GH' and 'IJ' are the two pairs. Shape-location pairings were randomized across participants, but consistent throughout the experiment for each participant.

The familiarization stimulus was a continuous sequence of pseudo-randomly ordered shape units. Units could not repeat and there were no breaks or delays between shapes or units such that TPs were 1.0 between shapes within units and .33 between shapes spanning unit boundaries. Test stimuli were 10, 2-shape sequences. Two sequences were pairs from the familiarization sequence (GH, IJ), two were embedded pairs (BC from 'ABC', EF from 'DEF'), and six were part-sequences composed of the last shape of one unit and the first shape of a different unit (e.g., FA).

Procedure. Participants sat 60 cm from the computer monitor. POG was calibrated using Experiment Builder software. Participants viewed the familiarization sequence for either 10 (80 repetitions of each unit presented) or 20 (160 repetitions of each unit) minutes, depending on their assigned condition. Participants were not given instructions other than to watch what appeared on the screen.

Following familiarization, participants completed a brief training session to familiarize them with a two-alternative forced-choice (2AFC) task. The training session consisted of 4 trials and employed the same procedure as the test phase, except that letters were presented rather than shapes. The test phase consisted of 12, 2AFC trials. In each trial, participants viewed two 2-shape sequences presented successively with a 750 ms pause between sequences. Participants were instructed to choose which was more familiar by clicking one of two corresponding mouse buttons. Half the test trials were "part vs. pair" trials that contrasted a part-sequence with a pair, and half were "part vs. embedded" trials that contrasted a part-sequence with an embedded pair. Part-sequences had no shapes in common with the pairs and embedded pairs against which they were contrasted. We presented test types in alternation, randomizing which we presented first and counterbalancing whether the part-sequence appeared first or second across trials. Table 3-1 provides a full example of the test sequences.

Table 3-1

Sample test sequences contrasted in Experiment 1

Part vs. Pair Contrasts		Part vs. Embedded Contrasts	
Part-Sequence	Pair	Part-Sequence	Embedded Pair
FA	GH	FA	BC
JD	GH	JD	BC
FI	GH	FI	BC
CD	IJ	CD	EF
HA	IJ	HA	EF
CG	IJ	CG	EF

Results and Discussion

Saccade latencies. Saccade latencies during familiarization were analyzed to assess

87

implicit learning of sequence structure. Latencies were calculated as the time from shape onset

until the initiation of the first eye movement that resulted in a fixation to that shape. Learning can

be inferred if mean saccade latency to the first shapes of units – whose locations are not

predictable from preceding shapes – are greater than mean saccade latency to the latter shapes of

units (second shape of pairs, second and third shapes of triplets) – whose locations are

predictable.

A 2 (familiarization duration: 10 vs. 20 mins.) x 2 (unit type: pair vs. triplet) x 2 (shape

number: first vs. latter) repeated-measures ANOVA revealed only a main effect of shape

number: $F(1,34) = 15.25$, $p < .001$, partial $\eta^2 = .31$; see Figure 3-2A. Saccade latencies were

significantly greater to the first shapes, relative to the latter shapes, of units, suggesting that

participants were sensitive to the unit structure of the sequence in both familiarization conditions.
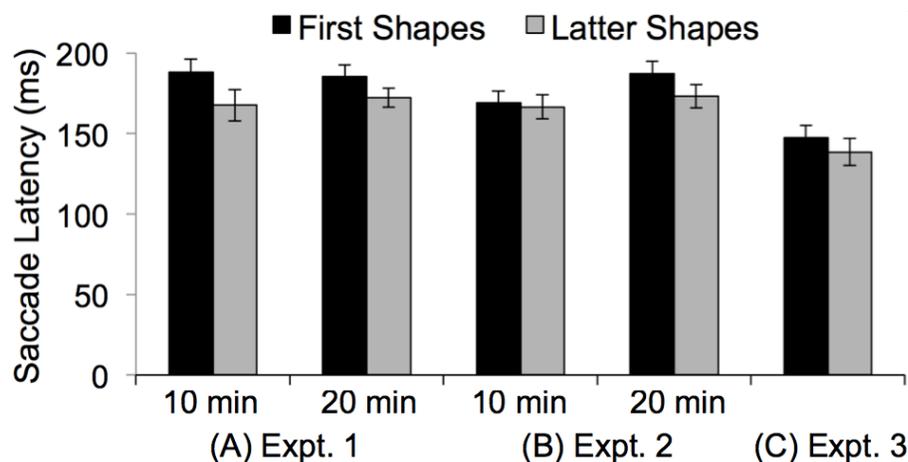
Figure 3-2. Mean saccade latency to the first and latter shapes of units in Experiments 1 (A) and 2 (B) by familiarization duration, and in Experiment 3 (C). Error bars represent standard error.

Button responses: Predictions. Both statistical and chunking models predict that

successful VSL should be indicated by participants' choosing pairs as more familiar than part-

sequences on part vs. pair trials. Statistical models also predict that participants should choose

(high TP) embedded pairs as more familiar than (low TP) part-sequences in part vs. embedded trials. In contrast, chunking models predict that participants may initially form chunks of all 2-shape sequences, but that representations of pairs and triplets will be strengthened as familiarization increases, even as representations of part-sequences and embedded pairs within triplets are weakened due to competition with units. Thus, chunking models predict that participants should fail to distinguish between embedded pairs and part-sequences, particularly after the longer (20 minute) familiarization (see Giroux & Rey, 2009).

Button responses: Results. Mean percentage of pair selections on part vs. pair trials, and embedded pair selections on part vs. embedded trials, were computed for the two familiarization conditions (Figure 3-3A). A 2 (test type) x 2 (familiarization duration) ANOVA revealed only a main effect of test type (F[1,34] = 8.61, p < .01, partial η^2 = .20). Participants chose pairs as more familiar than part-sequences more often then they chose embedded pairs as more familiar than part-sequences, regardless of familiarization condition. This finding may suggest that participants represented pairs more strongly than embedded pairs, as predicted by chunking models. Nevertheless, one-sample t-tests (this and all other t-tests reported were two-tailed) revealed that participants chose both pairs (t[35] = 9.21, p < .0001) and embedded pairs (t[35] = 4.93, p < .0001) as more familiar than part-pairs significantly more often than chance (50%), as predicted by statistical models.

Together these results do not clearly support either statistical or chunking models. However, because pairs and embedded pairs were not directly contrasted, it is difficult to draw strong conclusions about whether or not these sequences were represented differently. Experiment 2 was designed to address this issue. Test trials in Experiment 2 contrasted pairs and part-sequences as in Experiment 1, but also directly contrasted embedded pairs and pairs. If

89

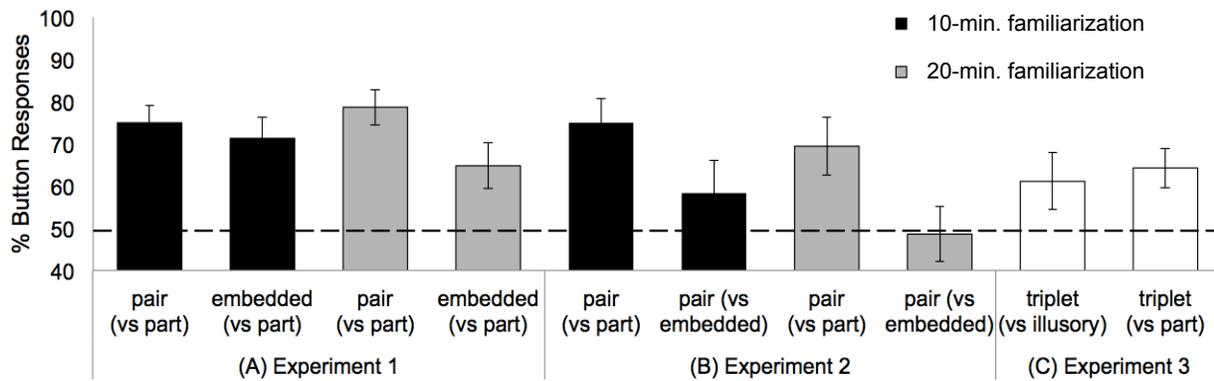*Figure 3-3.* Mean percentage of button responses for the various test types in Experiments 1 (A) and 2 (B) by familiarization duration condition, and in Experiment 3 (C). Error bars represent standard error. The dashed line indicates chance performance.

participants are equally familiar with pairs and embedded pairs, this suggests they primarily represent statistical relations between shapes. However, if participants choose pairs as more familiar than embedded pairs, this suggests participants represent some combination of both chunks and statistical relations (as embedded pairs were chosen as more familiar than part-sequences in Experiment 1).

Experiment 2

Method

Participants. Thirty-six undergraduate students were recruited and randomly assigned to a 10-minute ($N = 18$; 16 females; M age 20.6 years; range = 18.6 to 24.1) or 20-minute ($N = 18$; 13 females; M age 20.3 years; range = 19.0 to 22.2) familiarization condition, as in Experiment 1. Data from an additional 5 participants were excluded from the final sample due to sleepiness ($n = 4$) or failure to complete the experiment ($n = 1$).

Apparatus, stimuli, and procedure. The apparatus, stimuli, and procedure were identical to that of Experiment 1, with the following exceptions: (1) only two part-sequences

Table 3-2

Sample test sequences contrasted in Experiment 2

Part vs. Pair Contrasts		Pair vs. Embedded Contrasts	
Part-Sequence	Pair	Pair	Embedded Pair
FA	GH	GH	BC
CD	GH	GH	EF
FA	IJ	IJ	BC
CD	IJ	IJ	EF

were used: FA and CD, and (2) the test phase consisted of only 8, 2AFC trials. Half were "part vs. pair" trials, and half were "pair vs. embedded" trials (Table 3-2).

Results and Discussion

Saccade latencies. A 2 (familiarization duration) x 2 (unit type) x 2 (shape number) ANOVA revealed a main effect of shape number ($F[1,34] = 9.63$, $p < .01$, partial $\eta^2 = .22$), and interaction of shape number and exposure duration ($F[1,34] = 4.54$, $p < .05$, partial $\eta^2 = .12$); see Figure 3-2B. There were no other significant main effects or interactions. Post-hoc t-tests revealed that saccade latencies were significantly greater to the first shape of units in the 20-minute ($t[17] = 3.43$, $p < .01$), but not in the 10-minute ($t[17] = 0.75$, $p = .46$) familiarization condition. Saccade latencies to the first and latter shapes of units were not significantly different in the 10- and 20-minute conditions ($ts[34] < 1.71$, $ps > .05$).

These data suggest that participants were sensitive to the unit structure of the familiarization sequence after 20 minutes, but not 10 minutes, of exposure. It is unclear why this was the case, given that the familiarization phase was identical to that of Experiment 1, in which participants did show evidence of sensitivity to sequence structure after only 10 minutes. It could be that there was a ceiling effect among the participants in the 10-minute condition of

91

Experiment 2. Previous research suggests that it typically takes a minimum of 150 ms for an adult to program an eye movement (Fischer, Biscaldi, & Gezeck, 1997). Participants may have already been near ceiling, with saccade latencies to the first shapes of units being only 169 ms on average (see Figure 3-2B), such that they were unable to show significantly reduced saccade latencies to the latter shapes.

Button responses. Figure 3-3B shows the mean percentage of pair selections on part vs. pair and pair vs. embedded trials. A 2 (test type) x 2 (familiarization duration) ANOVA revealed only a main effect of test type ($F[1,34] = 10.53$, $p < .01$, partial $\eta^2 = .23$). Participants chose pairs as more familiar on significantly more trials when contrasted with part-sequences compared to when contrasted with embedded pairs, regardless of familiarization condition. Moreover, one-sample t-tests revealed that participants chose pairs as more familiar than part-sequences significantly more often than chance ($t[35] = 5.02$, $p < .0001$), but did *not* choose pairs as more familiar than embedded pairs significantly more often than chance ($t[35] = 0.68$, $p = .50$). These findings suggest participants represented pairs and embedded pairs similarly, as predicted by the statistical approach.

Overall, the results of Experiments 1 and 2 investigating adults' representation of embedded pairs in visual sequences suggest that participants represented statistical relations between items rather than chunks, a finding that contrasts with previous studies of embedded items conducted with auditory sequences (Giroux & Rey, 2009) and visual scenes (Fiser & Aslin, 2005). This difference is all the more striking given that our VSL task was designed to be as similar as possible to Giroux and Rey's auditory SL task.

It may be that learners represent both statistical relations and chunks (even proponents of the statistical approach argue that SL produces some kind of psychological units; e.g., Saffran,

2001), raising questions as to the relation between statistical and chunking processes (see Perruchet & Pacton, 2006). Another possibility, however, is that the assumption made by some chunking models – that higher-order chunks always compete with and replace lower-order chunks – may be incorrect. If learners were able under certain circumstances to maintain representations of various orders of chunks simultaneously, such as chunks and the smaller embedded chunks they contain, this might help to explain participants' performance in Experiments 1 and 2.

Servan-Schreiber and Anderson's (1990) 'competitive chunking' model proposes that (1) learners may be able to represent both lower-order chunks and the higher-order chunks that contain them, and (2) the familiarity of a sequence depends on the number of stored chunks needed to describe it. Thus, when participants viewed pairs and embedded pairs at test in Experiment 2, these sequences may have seemed equally familiar because pairs and embedded pairs were each represented by a single chunk, not because participants represented their underlying TPs. Similarly, when participants viewed part-sequences and embedded pairs in Experiment 1, embedded pairs may have seemed more familiar because they were represented by a single chunk whereas part-sequences were not, since their component shapes did not occur together consistently. The data from Experiments 1 and 2 cannot distinguish between these two interpretations – that learners represented statistical relations, or represented both embedded chunks and their larger (triplet) chunks.

Thus, Experiment 3 employed an illusory sequence design to: (1) examine how adults represent illusory visual sequences, and (2) distinguish between the statistical and competitive chunking explanations of Experiments 1 and 2. An illusory design can achieve this second aim because statistical and competitive chunking models make different predictions concerning the

93

fate of illusory items.

Experiment 3

Method

Participants. Eighteen undergraduate students ($N = 18$; 14 females; M age 19.6 years; range = 15.6 to 28.5) were recruited as in Experiments 1 and 2. Data from an additional 4 participants were excluded from the final sample due to poor calibration ($n = 1$), eye tracker failure ($n = 1$), or sleepiness ($n = 2$).

Stimuli. Stimuli were nine colored shapes that each loomed within one of 9 grid locations. Shapes were organized into six triplets, with each shape appearing in two triplets. The familiarization stimulus was a continuous sequence of pseudo-randomly ordered triplets. Triplets could repeat such that TPs were .50 between shapes within triplets and .33 between shapes spanning triplet boundaries. Triplets were organized such that two illusory triplets were created that had the same TP structure as triplets, but were never presented during familiarization (Figure 3-4). Hereafter, the nine shapes will be referred to by the letters ABCDEFGHI, where the six triplets are ABF, DBC, AEC, GHF, DHI, GEI, and the two illusory triplets are ABC and GHI.
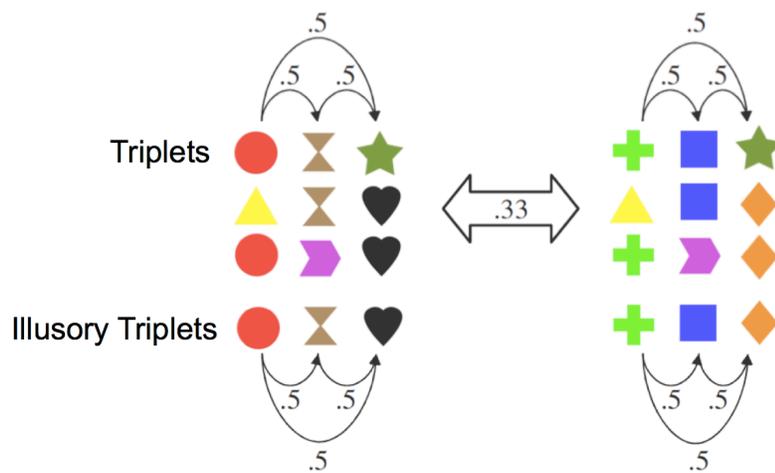

Figure 3-4. Sample triplets and illusory triplets used in Experiment. 3. Numbers above arrows indicate TPs between shapes.

94

Test stimuli were 10, 3-shape sequences. Six sequences were triplets, two were illusory triplets, and two were part-sequences composed of the last shape of one triplet and the first shape of a different triplet.

Apparatus and procedure. The apparatus and procedure were identical to Experiments 1 and 2, with the following exceptions: (1) participants viewed the familiarization sequence for 18 minutes (80 repetitions of each triplet presented), and (2) half of the test trials were "triplet vs. part," and half were "triplet vs. illusory" trials (Table 3-3).

Table 3-3

Sample test sequences contrasted in Experiment 3

Triplet vs. Part Contrasts		Triplet vs. Illusory Contrasts	
Triplet	Part-Sequence	Triplet	Illusory Triplet
ABF	IAE	ABF	ABC
DBC	FDH	DBC	ABC
AEC	IAE	AEC	ABC
GHF	FDH	GHF	GHI
DHI	FDH	DHI	GHI
GEI	IAE	GEI	GHI

Results and Discussion

Saccade latencies. Saccade latencies to the first and latter shapes of triplets were not significantly different: $t[17] = 0.94$, $p = .36$ (see Figure 3-2C). Thus, participants showed no oculomotor evidence of implicit learning of sequence structure. This was likely due to TPs between shapes within units being .50 in Experiment 3 (compared to 1.0 in Experiments 2 and 3), such that the latter shapes of units were not completely predictable from the previous shape, even if the triplet structure had been learned.

95

Button responses: Predictions. Both statistical and chunking models predict that successful VSL should results in triplets being more familiar than part-sequences. Statistical models also predict that triplets and statistically matched illusory triplets should seem equally familiar, whereas chunking models predict that triplets should seem more familiar because they are represented by a single higher-order chunk (e.g., 'ABF'), whereas illusory triplets are represented by two lower-order chunks (e.g., 'AB', 'BC'; Servan-Schreiber & Anderson, 1990) or no chunks at all (e.g., Fiser & Aslin, 2005). Even if illusory triplets are represented by two lower-order chunks, illusory triplets should seem relatively unfamiliar simply because a greater number of stored chunks are needed to describe them (Servan-Schreiber & Anderson, 1990).

Button responses: Results. A paired- samples t-test revealed that the percentage of trials on which participants chose triplets as more familiar (see Figure 3-3C) did not differ significantly when triplets were contrasted with illusory triplets versus part-sequences (t[17] = 0.24, p = .81). Moreover, one-sample t-tests revealed that participants chose triplets as more familiar than both part-sequences and illusory triplets significantly more often than chance: ts(17) > 2.81, ps < .02. These findings suggest that learners represent visual sequences in terms other than statistical relations between items, as predicted by chunking models.

General Discussion

The present series of experiments investigated processes of VSL. Specifically, we examined whether adults represent sequences in terms of chunks or statistical relations. We used highly comparable methods to examine performance on embedded and illusory item tasks that, in previous research, have suggested different underlying mechanisms.

Participants in Experiments 1 and 2 provided evidence of representing embedded pairs, contrary to the predictions of typical chunking models (e.g., Orbán et al., 2008), but consistent

with both statistical and competitive chunking models. Experiment 3 examined participants' endorsement of illusory items to distinguish between statistical and competitive chunking explanations. Participants distinguished triplets from statistically-matched illusory triplets, suggesting that they represented sequences in terms of chunks rather than statistics. Only the Servan-Schreiber and Anderson (1990) competitive chunking model is able to account for the data obtained across all three experiments.

Yet, the present data contrast with findings from previous studies of embedded (Fiser & Aslin, 2005; Giroux & Rey, 2009) and illusory (Endress & Mehler, 2009) items. This may mean that current models of SL are inadequate, as no single model can account for performance across tasks and domains. However, it is also possible that the representations resulting from SL are task-dependent such that representations vary depending on characteristics of the information to be learned. Adults may, for instance, represent units and their embedded chunks when the quantity of information or complexity of the task is relatively low (e.g., Experiments 1 and 2), but may represent only the highest order of chunks when a greater quantity or complexity of information puts additional demands on attention and memory systems (e.g., Fiser & Aslin, 2005). Further research is needed to examine these possibilities.

Regardless, the present experiments have important implications for theories and models of SL. Studies of chunking have a long history in the implicit learning literature, but have only recently been introduced to statistical learning research (Perruchet & Pacton, 2006). The present data suggest that our understanding of SL will profit from researchers continuing to consider the role chunking, particularly competitive chunking, may play in SL.

References

Elman, J. L. (1990). Finding structure in time. *Cognitive Science,* 14, 179-211.

Endress, A., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60,* 351-367.

Fischer, B., Biscaldi, M., & Gezeck, S. (1997). On the development of voluntary and reflexive components in human saccade generation. *Brain Research, 754,* 285-297.

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 higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning Memory and Cognition, 28,* 458-467.

Fiser, J., & Aslin, R. N. (2005). Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134,* 521.

Frank, M. C., Goldwater, S., Griffiths, T., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117,* 107-125.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33,* 260-272.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, 105,* 2745-2750.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10,* 233-238.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning.

Cognition, 81, 149-169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-month-old

infants. *Science, 274,* 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996a). Word segmentation: The role of

distributional cues. *Journal of Memory and Language, 35,* 606-621.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive

chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent.

Perception & Psychophysics, 67, 867-875.

PAPER 4:

When learning goes beyond statistics: Infants represent visual sequences in terms of chunks

Abstract

Much research has documented infants' sensitivity to statistical regularities in auditory and visual inputs, however the manner in which infants process and represent statistically defined information remains unclear. Two types of models have been proposed to account for this sensitivity: statistical models, which posit that learners represent statistical relations between elements in the input; and chunking models, which posit that learners represent statistically-coherent units of information from the input. Here, we evaluated the fit of these two types of models to behavioral data that we obtained from 8-month-old infants across two visual sequence-learning experiments. Experiments examined infants' representations of two types of structures about which statistical and chunking models make contrasting predictions: illusory sequences (Experiment 1) and embedded sequences (Experiment 2). In both experiments, infants discriminated between high probability sequences and low probability part-sequences, providing strong evidence of learning. Critically, infants also discriminated between high probability sequences and statistically-matched sequences (illusory sequences in Experiment 1, embedded sequences in Experiment 2), suggesting that infants represented statistically coherent chunks of elements from familiarization, rather than statistical relations between individual items. Together, these studies not only contribute substantial new data on infants' visual statistical learning ability, of which there is currently little, but also suggest that the representations that result from infants' visual statistical learning are best captured by chunking models.

When learning goes beyond statistics: Infants represent visual sequences in terms of chunks

How do learners make sense of their intricately structured auditory and visual environments? Previous research suggests that both infants and adults are able to identify statistically coherent pieces of information contained within larger sequences presented both aurally and visually (see Krogh, Vlach, & Johnson, 2012 for a review). This "statistical learning" ability may facilitate learners' detection of important types of environmental structure. For instance, statistical learning is thought to help learners identify words in continuous speech, facilitating language learning (e.g., Saffran, 2001), and help learners segment continuous motion into discrete events, facilitating visual learning and categorization (e.g., Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014).

Despite the scope and potential importance of statistical learning ability, the specific processes underlying statistical learning remain unclear. Recent research has investigated how two types of models of the mechanisms underlying statistical learning – statistical and chunking models – account for adults' statistical learning performances (see Thiessen, Kronstein, & Hufnagle, 2013 for a review). Although the data are not entirely conclusive, the extant data suggest that adults' statistical learning is best accounted for by chunking models (cf. Endress & Mehler, 2009). It remains unknown, however, which type of model best accounts for infants' statistical learning performances. This is an important issue to address, as statistical learning is posited to underlying much early learning, including language acquisition. Moreover, examining possible statistical and chunking processes in infants' statistical learning allows investigation of the extent to which the mechanisms underlying statistical learning are similar for infants and adults.

Statistical and Chunking Models of Statistical Learning

Two primary types of models have been developed to account for human statistical learning performance: statistical (or "transition-finding") models and chunking (or "clustering") models (Thiessen et al., 203). These models differ primarily in the representations they store in memory. Statistical models are typically instantiated with simple recurrent networks (SRNs) (e.g., Elman, 1990). These networks calculate and represent in memory statistical relations between items. For instance, one statistical relation that models (and human learners) may represent is transitional probability (TP). Transitional probability (TP) is defined as the probability of event Y given event X, and is a measure of the strength with which X predicts Y. Representing such a statistic would not only inform the model of the likelihood of two items occurring together, but would also allow the model to predict individual items based on previous items in a sequence. In this way, statistical models become sensitive to statistically coherent units of information contained within a sequence. Statistical models do not represent these units in memory, however; rather, they represent statistical relations between items.

Chunking models, in contrast, *do* represent statistically coherent units of information in memory. The mechanisms by which chunking models acquire these representations differ across models. Some of the most common chunking models are Bayesian models (e.g., Goldwater, Griffiths, & Johnson, 2006, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008) and PARSER (Perruchet & Vinter, 1998). Despite their varying learning processes, the representations that result from chunking models are discrete, statistically coherent, "chunks" of information. For instance, given some input, the Goldwater et al. (2009) Bayesian model formulates a set of hypotheses about the potential segmentations of the input, then assesses the likelihood of those segmentations. Of the segmentations that are consistent with the input, the segmentation with the

103

highest prior probability is the optimal solution, and the resulting segmented units are stored in memory. In contrast, PARSER (Perruchet & Vinter, 1998) implements basic laws of attention, memory, and associative learning. Items perceived within one attentional focus are "chunked" into a representational unit. Representations of units whose component items co-occur regularly will be progressively strengthened in memory, while representations of units whose component items do not co-occur regularly are forgotten.

Examining Model Fit to Human Data

Recently, research has investigated how well statistical and chunking models fit human data (e.g., Endress & Mehler, 2009; Fiser & Aslin, 2005; Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2015c). Many of these studies have examined the representations that adults store following auditory or visual statistical learning tasks, and whether these representations are best captured by statistical or chunking models. Such studies investigate representations of two types of items. The first type is illusory (or "phantom") units – units that are never presented to participants, but have the same statistical structure as other units that are presented. For example, if *tazepi*, *mizeru*, and *tanoru,* are words presented in a speech stream, and TPs are .50 between syllables within these words (e.g., between *ta* and *ze* and between *ze* and *ru*), a statistically matched illusory word would be *tazeru* (Endress & Mehler, 2009). Statistical models predict that units and illusory units should be indistinguishable because they are statistically equivalent. Chunking models, in contrast, predict that learners should fail to recognize illusory units because learners should not have extracted a chunk matching an illusory word from the input.

The second type of item researchers have investigated is embedded units – sub-units that occur only within larger units (Fiser & Aslin, 2005). In terms of linguistic materials, an

embedded item could be a group of syllables that occurs within a word, but never occurs independently (e.g., "eleph", as in "elephant") (Thiessen et al., 2012). Statistical models predict that, because learners represent statistical relations between all pairs of adjacent elements, as learners become familiar with a unit, distinguishing components embedded in that unit should improve relative to random configurations of elements. Many chunking models, in contrast, predict that as learners become familiar with a unit, they should become *less* able to distinguish components embedded in that unit from random configurations of elements (see Giroux & Rey, 2009). That is, representations of embedded units and their larger units compete in memory. Over time, economy of representation suggests that memory for units should be strengthened while competing representations of completely embedded units should weaken (Thiessen et al., 2013).

Five studies have recently investigated adults' representations of illusory and embedded units, and the ability of various models to account for this performance. Specifically, these studies have investigated adults' representations of illusory units presented in auditory sequences (Endress & Mehler, 2009) and visual sequences (Slone & Johnson, 2015c), and embedded units presented in auditory sequences (Giroux & Rey, 2009), visual sequences (Slone & Johnson, 2015c), and visual scenes (Fiser & Aslin, 2005; Orbán et al., 2008). Although the data are not entirely conclusive, the extant data suggests that adults' representations of these auditory and visual stimuli are best accounted for by chunking models (cf. Endress & Mehler, 2009).

It remains unknown, however, which type of model best accounts for infants' statistical learning performances. The two major types of chunking models (Bayesian models and PARSER) rely on assumptions about learners' priors (e.g. Goldwater et al., 2006, 2009) and attention, memory, and associative learning (Perruchet & Vinter, 1998) – factors that likely change between

infancy and adulthood. Thus, it is not clear whether or not the processes underlying statistical learning are similar for infants and adults. Although some studies acknowledge that infants and adults process information differently, they nevertheless assume that because infants and adults perform similarly in statistical learning experiments, these groups process information similarly (e.g., Endress & Mehler, 2009). Different types of processing (e.g., tracking statistical relations vs. storing chunks), however, can lead to similar performances on many statistical learning tasks.

The current set of experiments investigated whether statistical or chunking models best account for infants' statistical learning performances. Because most work comparing these models' fits to adult statistical learning data has employed visual stimuli, we focused on infants' visual statistical learning. Experiment 1 examined infants' learning of illusory visual sequences, and Experiment 2 examined infants' learning of embedded visual sequences. Both experiments asked: Are infants' representations of visual sequences best predicted by statistical models or chunking models?

Experiment 1

Participants. Sixteen healthy full-term 8-month-olds (13 females; *M* age 7 months 25 days, range = 6;30 to 8;17) were tested using visual familiarization. Data from an additional four infants were excluded from the final sample due to fussiness. Infants were recruited by letter and telephone from hospital records and given a small gift (a toy or baby t-shirt) for their participation.

Apparatus and stimuli. An Eyelink 1000 eye tracker (SR Research) with a 55.9-cm color monitor displayed stimuli and collected eye-tracking data.[1] A PC computer running

---

[1] Prior to familiarization, infants' point-of-gaze was calibrated using Experiment Builder software. Eye tracking data were recorded at 500 Hz during familiarization and test, but are not reported here as data quality was poor due primarily to the length of the study and infants' movement during familiarization.
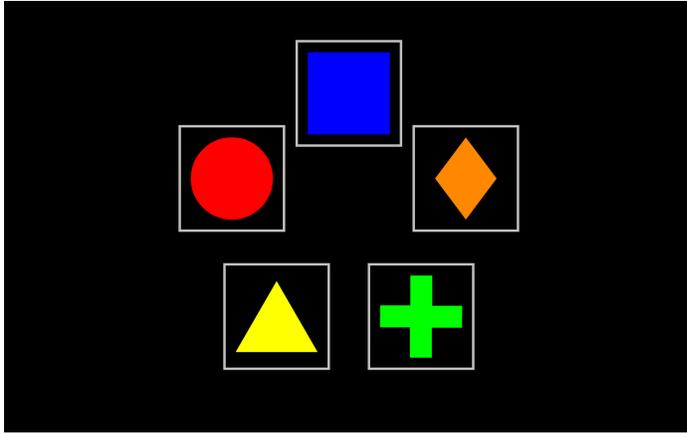
Figure 4-1. Sample spatial array and shapes presented in Experiments 1 and 2. Only one shape appeared at a time during familiarization and test.

Experiment Builder software controlled stimulus presentation and collected looking time data.

Stimuli consisted of a five-location spatial array and 5 colored shapes (Figure 4-1). Infants viewed a continuous sequence of shapes, presented one at a time for 750 ms each and looming from 1.5 to 5.5 cm in height (about 8.0°-10.0° visual angle) within one of the five locations on the eye-tracking monitor. The familiarization sequence consisted of a continuous sequence of repetitions of three units: two shape triplets (e.g., triplet 1: red circle, blue square, orange diamond; triplet 2: green plus, red circle, orange diamond), and one shape pair (e.g., blue square, yellow triangle). Units could repeat and there were no breaks or delays between units such that units were defined solely by distinctions in TPs. Because each shape could appear in more than one unit, TPs between shapes within units were either 1.0 or .50, and TPs between shapes spanning unit boundaries were .33 (Figure 4-2A).

Procedure. Infants sat on a caretakers' lap approximately 60 cm from the monitor. Caretakers were instructed not to interact with the infant or attend to the monitor. During familiarization and test, a trained observer who was unaware of the stimulus sequence viewed by the infant, observed the infant via a video feed from a camera placed directly below the monitor
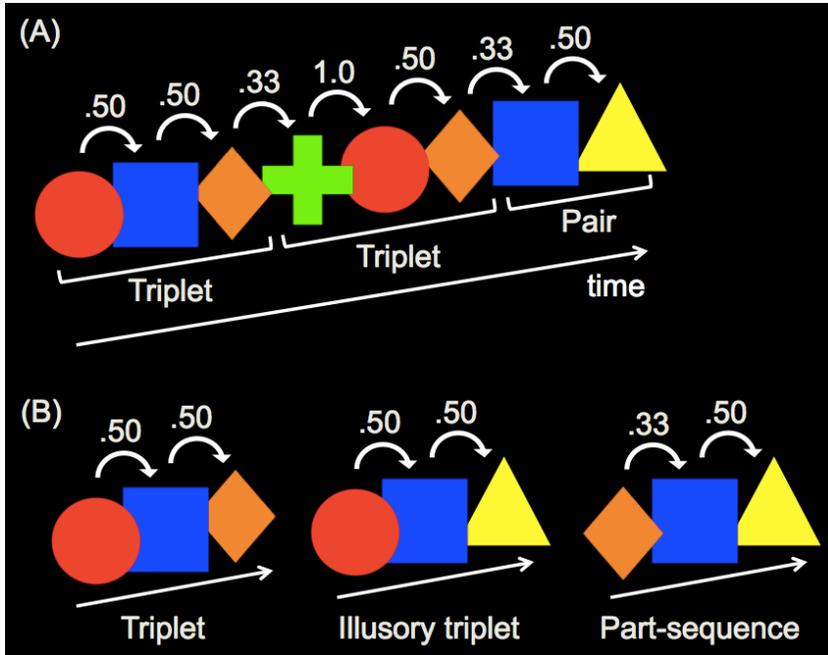
107

Figure 4-2. Sample (A) familiarization sequence and (B) test sequences presented in Experiment 1. Numbers above shapes represent TPs during familiarization. Brackets below shapes indicate the unit structure of the familiarization sequence.

and coded looking behavior online. Prior to familiarization, an attention-getter was presented to attract infants' attention to the center of the monitor. Familiarization began when the experimenter pressed a keyboard key to indicate that the infant was looking at the monitor. When the infant visually fixated away from the monitor, the experimenter released the key, causing the familiarization sequence to pause. If the infant returned attention to the monitor within 2 s, the familiarization sequence resumed. Otherwise, an attention-getter was shown to attract infants' gaze back to the monitor, whereupon the experimenter immediately resumed the familiarization sequence.

Familiarization continued until the full five-minute familiarization sequence (50 presentations of each unit) had been seen. Following familiarization, infants were presented with two blocks of three unique test trials, for a total of six test trials (see Figure 4-2B). Trial order

within each block was randomized. Each test trial consisted of a repetition of three shapes, with a 750 ms pause separating repetitions. Triplet test trials presented the triplet with .50 TPs between shapes during familiarization (e.g., red circle, blue square, orange diamond). Illusory triplet test trials presented a three-shape sequence that was never seen in its entirety during familiarization, but which had .50 TPs between adjacent shapes during familiarization (e.g., orange circle, blue square, yellow triangle). Thus, triplet and illusory triplets test sequences were equivalent in terms of TPs between adjacent shapes. Part-sequence test trials presented the last shape of a triplet followed by the pair from familiarization (e.g., orange diamond, blue square, yellow triangle), such that TP was .33 between the first two shapes and .50 between the latter two shapes. Thus, part-sequences had lower internal TPs compared to triplets and illusory triplets. An attention-getter was presented prior to the first test trial, and every subsequent trial, to center infants' gaze. Each test sequence was presented until the infant visually fixated away from the monitor for over 2 s or until 90 s had elapsed.

Coding and predicted results. Mean durations of looking to each of the three test types (triplet, illusory triplet, part-sequence) were computed. Looking time measures capitalize on the tendency for infants' general interest in a visual stimulus to decline upon repeated exposure over the familiarization period. When followed by presentation of test sequences that either bear resemblance to the original stimulus (i.e., triplet from familiarization), or differ along one or more dimensions (i.e., illusory triplet, part-sequence), significantly different durations of looking to these test types indicates discrimination between them, whereas equal durations of looking to the test types suggests no discrimination. The present study used a familiarization design, rather than an infant-controlled habituation design, to equate infants' exposure to the familiarization stimulus prior to test. One limitation of using familiarization rather than habituation, however, is

109

that we cannot accurately predict a priori whether infants will exhibit a familiarity or a novelty preference as an index of learning (see Hunter & Ames, 1988). Thus, while we predicted discrimination between test types, we did not have a priori predictions about the directions of infants' preferences.

Both statistical and chunking models predict that successful learning should be indicated by significantly different durations of looking on triplet (high TP) compared to part-sequence (lower TP) test trials. Statistical and chunking models differ, however, in their predictions about infants' relative looking durations on illusory triplet test trials. Statistical models predict that infants should show *similar* durations of looking to triplet and illusory triplet test sequences because these sequences had equivalent TPs between adjacent shapes during familiarization (TPs = .50). In contrast, chunking models predict that infants should show *different* durations of looking to triplet and illusory triplet test sequences. That is, chunking models posit that, with enough exposure, learners should represent the highest-level units most strongly in memory. Thus, illusory triplets should seem relatively novel to infants because these units were never presented in their entirety during learning.

Results and Discussion

Preliminary analyses revealed that the distributions of infants' looking times to the three test types were somewhat positively skewed (M skewness = 0.81, SD = 0.63) and not normally distributed (M kurtosis = 1.53, SD = 0.86). Thus, we calculated log transformations of each infants' mean looking duration to each test type. These log-transformed measures were less skewed (M skewness = -0.19, SD = 0.23) and more normally distributed (M kurtosis = -0.95, SD = 0.05) than were infants' raw looking times, and were therefore used for all further analyses (Figure 4-3 shows raw looking times for ease of interpretation).
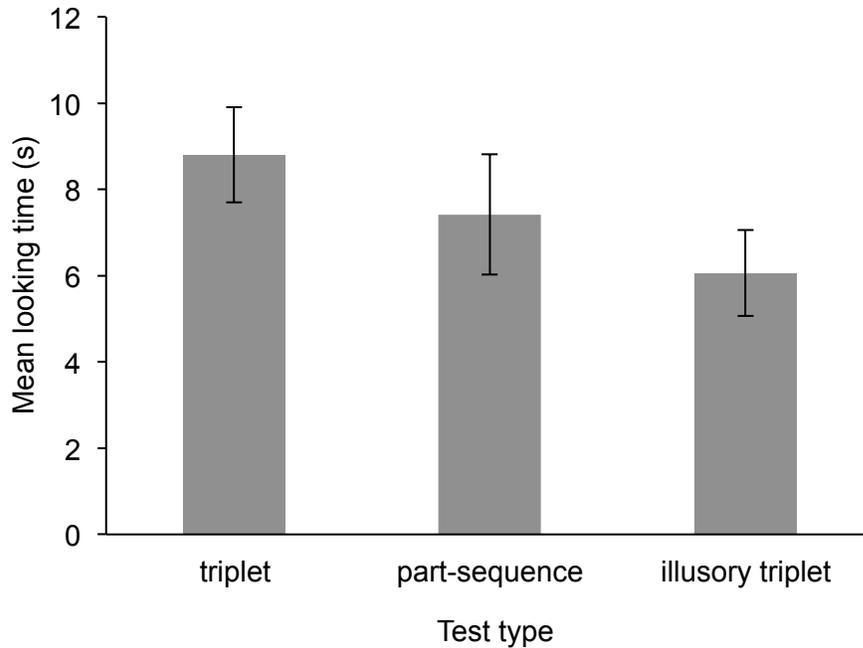
Figure 4-3. Infants' mean looking duration to each of the three test types in Experiment 1.

A repeated measures analysis of variance (ANOVA) revealed a main effect of test type: $F(1,15) = 12.53$, $p < .01$, partial $\eta^2 = .455$. Planned paired samples t-tests (this and all subsequent t-tests were two-tailed) revealed that infants looked significantly longer during triplet test trials, compared to both part and illusory triplet test trials: ts$(16) > 2.69$, ps$ < .02$, ds$ > 0.66$. Looking duration did not differ significantly on part and illusory triplet test trials: $t(15) = 2.05$, $p = .06$, $d = 0.51$ (Figure 4-3).

Infants were able to discriminate triplets from both part-sequences and illusory triplets. In contrast to the predictions of statistical models, infants did not represent the familiarization sequence primarily in terms of TPs between adjacent items. Rather, infants discriminated between test sequences containing equal TPs (that is, between triplets and illusory triplets), suggesting that infants represented the familiarization sequence in terms of extracted units, not statistical relations.

The above interpretation prioritizes relations between adjacent items on account of data suggesting that non-adjacent relations are more difficult for learners to identify compared to adjacent relations (Newport & Aslin, 2004). It is possible, however, that a statistical model allowing for computation of non-adjacent relations could also account for the data from Experiment 1. For instance, both triplet and illusory triplet test sequences had .50 TPs between adjacent shapes during familiarization. Nevertheless, triplet test sequences also had .50 TP between non-adjacent shapes during familiarization (i.e., red circle predicted orange diamond appearing two items later with .50 accuracy), whereas illusory triplet test sequences had .00 TP between non-adjacent shapes during familiarization (i.e., red circle never predicted yellow triangle appearing two items later). Additionally, part-sequences had .33 TP between non-adjacent shapes during familiarization (i.e., orange diamond predicted yellow triangle appearing two items later with .33 accuracy). It is possible, therefore, that infants' discrimination between triplet test trials and both illusory and part-sequence test trials could have resulted from infants' representing statistical relations between non-adjacent shapes (either independently or in addition to representing relations between adjacent shapes), rather than representing chunks.

Experiment 2 was designed to address this possibility by examining infants' representations of embedded sequences. In contrast to illusory sequences, embedded sequences can consist of only two items, such that only adjacent relations can be used to distinguish embedded sequences from other test sequences. If infants discriminate such embedded pairs from statistically matched non-embedded pairs in Experiment 2, this is further evidence in favor of chunking, rather than statistical models. If, however, without added non-adjacent statistical cues, infants do not discriminate embedded pairs from statistically matched non-embedded pairs,

112

this suggests that performance in Experiment 1 may in fact have resulted from representations of non-adjacent statistical relations.

Experiment 2

Method

  Participants. Eighteen healthy full-term 8-month-olds (5 females; *M* age 8 months 9 days, range = 7;7 to 8;27) were recruited as in Experiment 1. Data from an additional 10 infants were excluded from the final sample due to fussiness (*n* = 6), parental interference (*n* = 2), failure to look at the monitor (*n* = 1), or excessive squirming (*n* = 1).

  Apparatus and stimuli. The apparatus and stimuli were identical to those of Experiment 1, except for the unit structure of the familiarization sequence. In Experiment 2, the familiarization sequence consisted of a continuous sequence of repetitions of two units: a shape triplet (e.g. yellow triangle, blue square, orange diamond), and a shape pair (e.g. green plus, red circle). Units could repeat and there were no breaks or delays between units such that units were defined solely by distinctions in TPs. Because each shape appeared in only one unit, TPs between shapes within units were 1.0, and TPs between shapes spanning unit boundaries were .50 (Figure 4-4A).

  Procedure. The procedure was identical to that of Experiment 1, with two exceptions. First, because only two units were presented in Experiment 2 (compared to three units in Experiment 1), infants saw 80 presentations of each unit (compared to 50 in Experiment 1) during the five-minute familiarization. Second, different test types were presented (Figure 4-4B). Pair test trials presented the pair from familiarization (e.g., green plus, red circle; TP=1.0), embedded pair test trials presented the latter two shapes of the triplet from familiarization (e.g., blue square, orange diamond; TP=1.0), and part-sequence test trials presented the last shape of
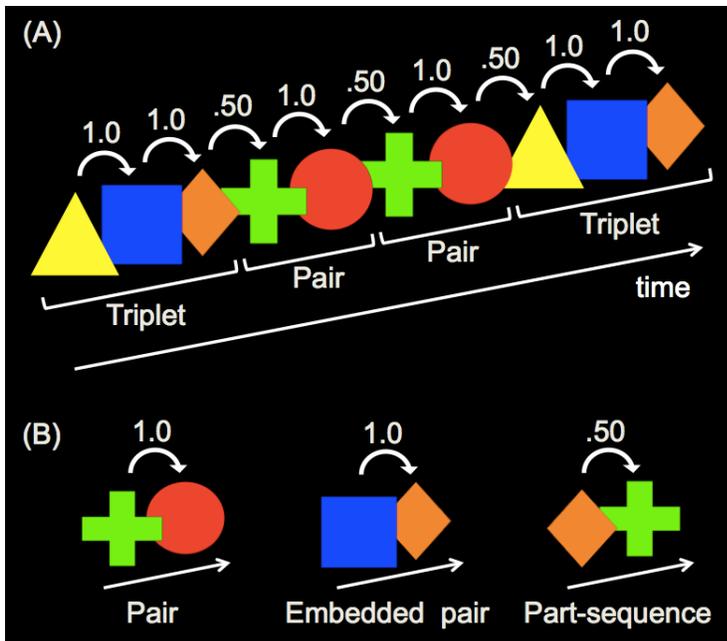
113

Figure 4-4. Sample (A) familiarization sequence and (B) test sequences presented in Experiment 2. Numbers above shapes represent TPs during familiarization. Brackets below shapes indicate the unit structure of the familiarization sequence.

the triplet followed by the first shape of the pair from familiarization (e.g., orange diamond, green plus; TP=.50). Thus, pairs and embedded pairs had equal TPs between shapes, and higher TPs than part-sequences.

Predicted results. Both statistical and chunking models predict that successful learning should be indicated by significantly different durations of looking on pair (high TP) compared to part-sequence (lower TP) test trials. Statistical and chunking models differ, however, in their predictions about infants' relative looking durations on embedded pair test trials. Statistical models predict that infants should show *similar* durations of looking to pair and embedded pair test sequences because these sequences had equivalent TPs between shapes during familiarization (TP=1.0). In contrast, chunking models posit that representations of embedded pairs compete with representations of their larger triplets in memory such that embedded pairs may be represented relatively weakly compared to pairs (which are not embedded and therefore

114

do not compete with encompassing representations), resulting in *different* durations of looking to pair versus embedded pair test sequences.

Results and Discussion

Preliminary analyses revealed that the distributions of infants' looking times to all three test types were positively skewed (*M* skewness = 2.27, *SD* = 0.46) and not normally distributed (*M* kurtosis = 5.79, *SD* = 2.63). Thus, as in Experiment 1, we calculated log transformations of each infants' mean looking duration to each test type. These log-transformed measures were less skewed (*M* skewness = 0.44, *SD* = 0.28) and more normally distributed (*M* kurtosis = -0.37, *SD* = -0.37) than were infants' raw looking times, and were therefore used for all further analyses (Figure 4-5 shows raw looking times for ease of interpretation).

A repeated measures analysis of variance (ANOVA) revealed a main effect of test type: $F(1,17) = 5.00, p < .05$, partial $\eta^2 = .227$. Planned paired samples *t*-tests (this and all subsequent *t*-tests were two-tailed) revealed that infants looked significantly longer during both part-
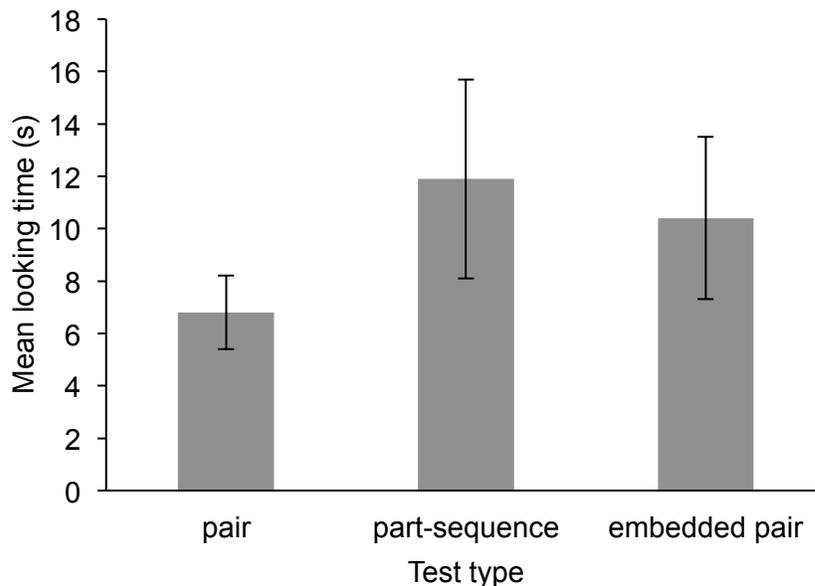

Figure 4-5. Infants' mean looking duration to each of the three test types in Experiment 2.

sequence and embedded test trials, compared to pair test trials: ts(17) > 2.17, ps < .05, ds > 0.51. Looking duration did not differ significantly on embedded compared to part-sequence test trials: t(17) = 0.44, p = .67, d = 0.10 (Figure 4-5).

Infants were able to discriminate pairs from both part-sequences and embedded pairs. In contrast to the predictions of statistical models, infants did not represent the familiarization sequence primarily in terms of TPs between adjacent items. Rather, infants discriminated between test pairs containing equal TPs (that is, between pairs and embedded pairs), suggesting that infants represented non-embedded sequences differently (and likely more strongly) than embedded sequences, supporting the predictions of chunking models. Moreover, infants' ability to discriminate sequences with equivalent statistical relations between adjacent items and no additional non-adjacent statistical cues, suggests that performance in Experiment 1 may also have been due to infants' representations of chunks from familiarization, rather than non-adjacent statistical relations.

General Discussion

Investigations of model fit to human data are a necessary part of the effort to characterize human learning (Frank et al., 2010). The present experiments examined infants' learning of two types of structures about which statistical and chunking models make contrasting predictions: illusory items (Experiment 1) and embedded items (Experiment 2). In both experiments, infants discriminated between units (i.e. triplets in Experiment 1, pairs in Experiment 2) and part-sequences, providing strong evidence of learning. Infants also discriminated units from statistically matched illusory triplets (Experiment 1) and embedded pairs (Experiment 2), suggesting that infants represented statistically coherent chunks from familiarization, rather than statistical relations between individual items. Together, these studies not only contribute

116

substantial new data on infants' visual statistical learning ability, of which there is currently little, but also suggest that the representations that result from infants' visual statistical learning are best captured by chunking models.

As mentioned previously, the present experiments used familiarization, rather than infant-controlled habituation, to equate infants' exposure to the familiarization stimulus prior to test. Exposure was equated within each experiment because chunking models typically assume that the relative representational strengths of smaller units (e.g., embedded units) compared to their larger encompassing units changes over time, due to assumptions of economy of representation and competition between representations (Thiessen et al., 2013). Using familiarization rather than infant controlled habituation therefore helped to ensure that infants had: (1) substantial exposure to the familiarization stimuli (i.e., five minutes is much longer than infants typically expose themselves to these types of stimuli in habituation studies), and (2) equal amounts of time to form representations of the familiarization stimuli.

Nevertheless, one limitation of using familiarization rather than habituation is that we cannot accurately predict a priori whether infants will exhibit familiarity or novelty preferences as indices of learning (see Hunter & Ames, 1988). In both experiments, infants discriminated units from part-sequences, indicating that they learned the structure of the familiarization sequence, however infants showed different directions of preference in Experiments 1 and 2. The preference for part-sequences in Experiment 2 follows the pattern of many infant statistical learning experiments (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Slone & Johnson, 2015a; Thiessen, Hill, & Saffran, 2005, Experiment 2) and is typically interpreted as a novelty preference for sequences that were not as strongly represented during the learning phase. The preference for units (triplets) in Experiment 1 has also been demonstrated in

117

some infant statistical learning experiments (e.g., Fiser & Aslin, 2002; Graf Estes, 2012, Experiment 1; Thiessen et al., 2005, Experiment 1), and is consistent with the notion that direction of preference depends on the complexity of the task, with a greater tendency to find familiarity preferences as task complexity increases (Hunter and Ames, 1988).

Several differences between the familiarization sequences in Experiments 1 and 2 likely increased the relative complexity and difficulty of Experiment 1, resulting in infants' familiarity preference in this study. First, the familiarization sequence consisted of three units (two triplets and one pair) in Experiment 1 but only two units (one triplet and one pair) in Experiment 2. Second, shapes could appear in more than one unit in Experiment 1. This resulted in lower TPs between shapes within units in Experiment 1 (.50 TPs) compared with Experiment 2 (1.0 TPs), as well as smaller differences in TPs between adjacent items within units and adjacent items within part-sequences in Experiment 1 (triplets: .50-.50 TPs; part-sequences: .33-.50 TPs; see Figure 2B) compared with Experiment 2 (pair: 1.0 TPs; part-sequences: .50 TPs; see Figure 4B). Additionally, although infants viewed the familiarization sequence for the same duration (5 minutes) in both experiments, infants saw a greater number of presentations of each unit in Experiment 2 (80 presentations of each unit) compared to Experiment 1 (50 presentations of each unit). Greater exposure to each unit in Experiment 2 may have resulted in a novelty preference due to habituation, whereas less exposure to each unit in Experiment 1 may have led infants to look longer to sequences more reminiscent of the familiarization stimuli.

Comparison to Previous Literature

The present experiments add to a growing body of literature suggesting that infants are able to detect statistically coherent sequences of information contained within larger visual sequences (e.g., Kirkham, Slemmer, & Johnson, 2002; Marcovitch & Lewkowicz, 2009; Slone &

118

Johnson, 2015a; Stahl et al., 2014). These studies have primarily examined infants' detection of pairs of shapes within a continuous sequence of shapes. The present findings contrast with a recent study suggesting 8-month-olds' inability to represent the structure of visual sequences composed of triplets (Slone & Johnson, 2015b). It is possible that increased exposure to the learning sequence in the present experiments compared to those of Slone and Johnson (2015b) facilitated infants' learning. It is also possible that the present familiarization structures were better learned because they consisted of both pairs and triplets, rather than only triplets as in Slone and Johnson (2015b). If infants are better at detecting units consisting of two rather than three shapes, infants' learning of the pairs in the present experiments may have facilitated infants' detection of the triplets. Additionally, the present experiments also contained spatial cues to sequence structure, whereas the Slone and Johnson (2015b) experiments presented each shape in the same central location. Future research could examine which of these factors may have facilitated learning in the present experiments by varying the length of familiarization, whether infants are exposed to triplets only or both pairs and triplets, and the spatial locations of shapes. Regardless, however, the present findings suggest that infants are able to detect triplet structures in visual sequences, given the right stimuli (see also Stahl et al., 2014).

Experiments 1 and 2 are the first to examine infants' representations of illusory and embedded items and whether statistical or chunking models best capture those representations. That infants' performances align most closely with the predictions of chunking models is consistent with many previous findings with adults (Fiser & Aslin, 2005; Giroux & Rey, 2009; Slone & Johnson, 2015c). Infants' failure to distinguish pairs from embedded pairs in Experiment 2, however, contrasts with the findings of Slone and Johnson (2015c, Experiment 2), in which adults failed to distinguish between visual pair sequences and embedded pair

119

sequences. The familiarization sequence used in the present study was highly similar to that of Slone and Johnson (2015c), except for the number of units employed (one triplet and one pair in the present study's Experiment 2; two triplets and two pairs in Slone & Johnson, 2015c). Moreover, infants' and adults' different performances across these two studies cannot be attributed to the number of presentations of each unit (80 presentations in both studies). Slone & Johnson (2015c) interpreted their results as evidence that adults maintained representations of both lower-order embedded pairs as well as the larger triplets that contained them, as has been suggested by at least one chunking model (Servan-Schreiber & Anderson, 1990). If this interpretation is correct, the present findings could suggest that infants processed visual sequences differently from adults, representing only the highest order units, rather than also representing lower-order embedded units. Future research should investigate this possibility to examine whether there are in fact differences in the representations that infants and adults generate via statistical learning.

Future research should also examine the representations that infants acquire from statistical learning of auditory sequences, and whether these are best reflected by statistical or chunking models. Some findings from auditory statistical learning studies suggest that infants treat the representations resulting from statistical learning as actual linguistic items (Saffran, 2001), mapping them as labels for objects (e.g., Graf Estes, Evans, Alibali, & Saffran, 2007). These findings seem to indicate that infants may represent units of information from auditory statistical learning tasks, rather than statistical relations between items, as suggested by chunking models. However, past research with adults, and recent research with infants, suggests that statistical learning may operate differently across modalities (e.g., Conway & Christiansen, 2005; Emberson, Conway, & Christiansen, 2011; Saffran, 2002; Slone & Johnson, 2015b).

120

Examining infants' representations of illusory and embedded auditory items could help to determine whether learners represent statistical information similarly across modalities in infancy. Moreover, auditory statistical learning methods may provide additional information about infants' representations than can be generated with visual methods. Auditory sequences are typically presented more rapidly than visual sequences, and infants can listen to auditory sequences while moving about, making it is easier to expose infants to greater amounts of auditory compared to visual information. This may allow researchers to examine chunking models' predictions about how the relative representational strengths of embedded and non-embedded units changes with increased exposure.

References

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9, 321-324. DOI:10.1111/1467-9280.00063

Conway, C. M. and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24-39. DOI:10.1037/0278-7393.31.1.24

Elman, J. L. (1990). Finding structure in time. *Cognitive Science,* 14, 179-211. DOI:10.1207/s15516709cog1402_1

Emberson, L. L., Conway, C. M., and Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology* 64, 1021-1040. DOI:10.1080/17470218.2010.538972

Endress, A. and Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351-367. DOI: 10.1016/j.jml.2008.10.003

Hunter, M. and Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*, *5*, 69-95. DOI

Fiser, J. and Aslin, R. N. (2005). Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134,* 521. DOI: 10.1037/0096-3445.134.4.521

Frank, M. C., Goldwater, S., Griffiths, T., and Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117,* 107-125. DOI:

10.1016/j.cognition.2010.07.005

Giroux, I. and Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33,* 260-272. DOI: 10.1111/j.1551-6709.2009.01012.x

Goldwater, S., Griffiths, T. L., and Johnson, M. (2006). Contextual dependencies in unsupervised word segmentation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics* Association for Computational Linguistics, Sydney, Australia, pp. 673-680. DOI: 10.3115/1220175.1220260

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21-54. DOI: 10.1016/j.cognition.2009.03.008

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in psychology*, *3*. DOI: 10.3389/fpsyg.2012.00447

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*, 254-260. DOI:10.1111/j.1467-9280.2007.01885.x

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42. DOI:10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., and Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78,* 1559-1571. DOI:10.1111/j.1467-8624.2007.01083.x

Krogh, L., Vlach, H. A., and Johnson, S. P. (2012). Statistical learning across development:

flexible yet constrained. *Frontiers in psychology*, *3,* 598. DOI: 10.3389/fpsyg.2012.00598

Marcovitch, S. and Lewkowicz, D. L. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science, 12,* 1020-1025. DOI:10.1111/j.1467-7687.2009.00838.x

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127-162. DOI: 10.1016/S0010-0285(03)00128-2

Orbán, G., Fiser, J., Aslin, R. N., and Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, 105,* 2745-2750. DOI: 10.1073/pnas.0708424105

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263. DOI: 10.1006/jmla.1998.2576

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*, 149-169. DOI: 10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172-196. DOI:10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928. DOI:10.1126/science.274.5294.1926

Servan-Schreiber, E. and Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592. DOI: 10.1037/0278-7393.16.4.592

Slone, L. K. and Johnson, S. P. (2015a). Infants' statistical learning: 2-and 5-month-olds' segmentation of continuous visual sequences. *Journal of experimental child psychology*,

133, 47-56. DOI: 10.1016/j.jecp.2015.01.007

Slone, L. K. and Johnson, S. P. (2015b). *Modality constraints on statistical learning in infancy*. Manuscript submitted for publication.

Slone, L. K. and Johnson, S. P. (2015c). Statistical and chunking processes in adults' visual sequence learning. In R. P. Cooper (Ed.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Pasadena, CA: Cognitive Science Society.

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., and Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development, 85,* 1821-1826. DOI: 10.1111/cdev.12247

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7,* 53-71. DOI:10.1207/s15327078in0701_5

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792. DOI: 10.1037/a0030801