Statistical and Chunking Processes in Adults' Visual Sequence Learning

Lauren K. Slone (laurenkslone@gmail.com)
Department of Psychology, 1285 Franz Hall
Los Angeles, CA 90095-1563 USA

Scott P. Johnson (scott.johnson@ucla.edu)
Department of Psychology, 1285 Franz Hall
Los Angeles, CA 90095-1563 USA

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.

Keywords: statistical learning; transitional probability; chunking; implicit learning

Introduction

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 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 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.

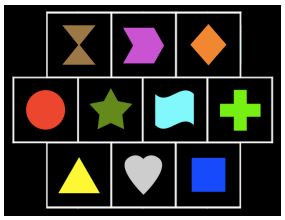

Figure 1: Sample stimulus array used in Experiment 1. Only one shape appeared on the screen at a time.

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 1 provides a full example of the test sequences.

Table 1: Sample test sequences contrasted in Expt. 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 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.

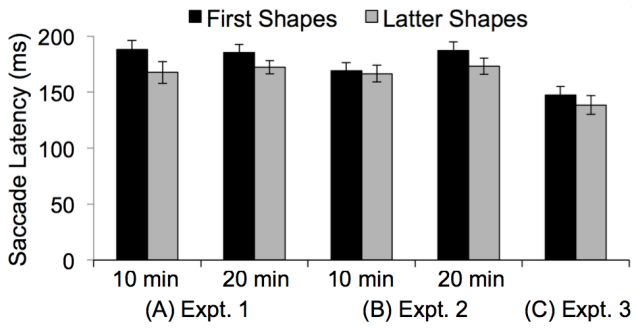

Figure 2. Mean saccade latency to the first and latter shapes of units in Expts 1 (A) and 2 (B) by familiarization duration, and in Expt 3 (C). Error bars represent standard error.

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 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.

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 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 $\eta^2 = .20$). 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 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

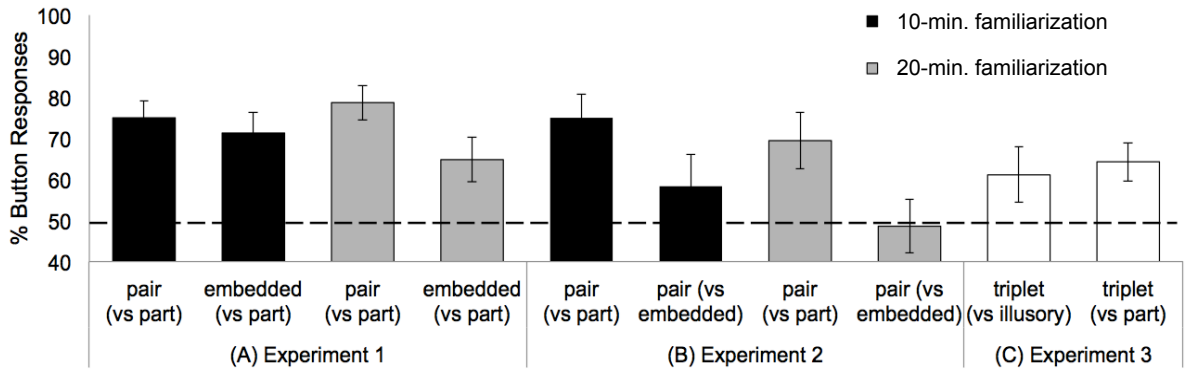

Figure 3. Mean percentage of button responses for the various test types in Experiments 1 (A) and 2 (B) by familiarization duration condition, and in Expt. 3 (C). Error bars represent standard error. The dashed line indicates chance performance.

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 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 2).

Table 2: Sample test sequences contrasted in Expt. 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

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 η^2 = .22), and interaction of shape number and exposure duration (F[1,34] = 4.54, p < .05, partial η^2 = .12); see Figure 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 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 2B), such that they were unable to show significantly reduced saccade latencies to the latter shapes.

Button Responses Figure 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 η^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 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 9 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 4). Hereafter, the 9 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.

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.

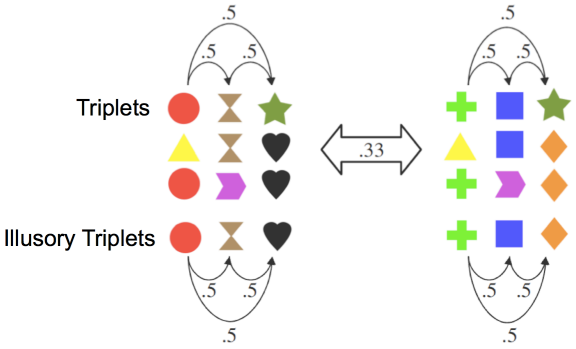

Figure 4. Sample triplets and illusory triplets used in Expt. 3. Numbers above arrows indicate TPs between shapes.

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).

Table 3: Sample test sequences contrasted in Expt. 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 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.

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 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 of chunks 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.

Acknowledgments

This research was funded by NIH R01-HD73535 to SPJ and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-0707424 to LKS. The authors would like to thank the participants and the volunteers and staff of the UCLA Baby Lab.

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.