CrossMark

The roles of order, distance, and interstitial items in temporal visual statistical learning

Su Hyoun Park[1] · Leeland L. Rogers[1] · Timothy J. Vickery[1]

Abstract

Humans are adept at learning regularities in a visual environment, even without explicit cues to structure and in the absence of instruction—this has been termed "visual statistical learning" (VSL). The nature of the representations resulting from VSL are still poorly understood. In five experiments, we examined the specificity of temporal VSL representations. In Experiments 1A, 1B, and 2, we compared recognition rates of triplets and all embedded pairs to chance. Robust learning of all structures was evident, and even pairs of non-adjacent items in a sequentially presented triplet (AC extracted from a triplet composed of ABC) were recognized at above-chance levels. In Experiment 3, we asked whether people could recognize rearranged pairs to examine the flexibility of learned representations. Recognition of all possible orders of target triplets and pairs was significantly higher than chance, and there were no differences between canonical orderings and their corresponding randomized orderings, suggesting that learners were not dependent upon originally experienced stimulus orderings to recognize co-occurrence. Experiment 4 demonstrates the essential role of an interstitial item in VSL representations. By comparing the learning of quadruplet sets (e.g., ABCD) and triplet sets (e.g., ABC), we found learning of AC and BD in ABCD (quadruplet) sets were better than the learning of AC in ABC (triplet) sets. This pattern of results might result from the critical role of interstitial items in statistical learning. In short, our work supports the idea of generalized representation in VSL and provides evidence about how this representation is structured.

Keywords Visual statistical learning · Generalization · Associative learning

Introduction

Humans inhabit a dynamic world where objects are constantly appearing, disappearing, and moving through space and time. Nevertheless, we do not struggle to process this massive influx of data. One potential coping mechanism is to learn and exploit statistic regularities, which are replete in human perceptual environments. Indeed, many empirical studies suggest that humans are sensitive to statistical regularities in experimental contexts, and can use such regularity to improve performance (e.g., Chun, 2000, 2003; Chun & Jiang, 1998; Turk-Browne, Jungé, & Scholl, 2005), and to build novel representations by binding statistically associated items together (Fiser & Aslin, 2001, 2002, 2005; Glicksohn & Cohen, 2011; Turk-Browne & Scholl, 2009). One such phenomenon, called

✉ Timothy J. Vickery
  tvickery@psych.udel.edu

[1] Department of Psychological and Brain Sciences, University of Delaware, 108 Wolf Hall, Newark, DE 19716, USA

visual statistical learning (VSL), refers to the visual system's ability to detect, extract, and later recognize regularities within an environment (Brady & Oliva, 2008; Fiser & Aslin, 2001, 2002, 2005; Otsuka, Nishiyama, Nakahara, & Kawaguchi, 2013; Turk-Browne, 2012; Turk-Browne et al., 2005). It has been discovered that humans have an ability to learn regularities in temporal (e.g., Fiser & Aslin, 2002; Turk-Browne et al., 2005) and spatial contexts (Chun, 2000, 2003; Fiser & Aslin, 2001; Oliva & Torralba, 2007).

In many "temporal" VSL experiments, subjects experience a continuous temporal stream of novel shapes in which (unbeknownst to participants) the shapes appear according to a temporal structure, such that one shape will always follow another shape (e.g., if A, B and C are a triplet, A always precedes B and B always precedes C in the training sequence). While these structured sets are repeated several times in the familiarization phase of the experiment, participants are usually unaware of the contingencies among shape appearances, both because the experimenter withholds any information about the underlying structure of the shapes and because the structured sets are presented in a pseudo-randomized order

(e.g., D-E-F-A-B-C-G-H-I, etc.). Following the familiarization task, participants perform a familiarity judgment task by using two-alternative forced choice (2AFC) tests to measure whether participants have learned the visual regularities (Fiser & Aslin, 2001, 2002, 2005; Turk-Browne, 2012). Under many different circumstances, it has been demonstrated that participants do learn such regularities.

While it is known that statistical regularities can be learned under different conditions, uncertainty surrounds the form of the resulting representations. Fiser and Aslin (2002) sought to characterize these representations. During a forced-choice familiarity discrimination phase, subjects were forced to choose between the base triplet target set (e.g., ABC) and a rearranged triplet set that included one embedded pair in the base triplet target at test (e.g., CAB or BCA). Only the part triplet set whose first shape pair was not an embedded pair (i.e., CAB) was discriminable from the base triplet, suggesting that order was important to recognizing contingent sets, with recognition privileging the first possible contingency exposed at test. This work provided insight into how representations created from VSL are expressed: the temporal order of stimuli within a structured group appears to be important for recognition.

Other studies showed that the nature of the representation produced by VSL is contingent on a hierarchical structure such that items with regularities eventually become grouped into a larger unit (e.g., Fiser & Aslin, 2005; Glicksohn & Cohen, 2011). In other words, when there is a larger complex feature (e.g., triplet of ABC), VSL does not allow for the learning of embedded sub-features (e.g., AB, BC, and AC). Fiser and Aslin (2005) asked how this feature hierarchy was represented in VSL; they sought to determine whether all levels (i.e., triplet and its embedded pairs) would be equally represented or only the highest one (i.e., triplet) would be represented (Fiser & Aslin, 2005). After the familiarization phase, participants had to choose the more familiar set between base triplets (or pairs) and foil triplets (or pairs). The results showed that participants could discriminate base triplets from foil triplets, which means that they could represent the highest-level features, but they could not discriminate between embedded pairs and foil pairs. These results suggested there is a hierarchical structure in VSL, such that once triplets were learned as a group in the context of VSL, it became prioritized over learning any sub-features of that base triplet (i.e., embedded pairs). Although humans rely heavily on visual statistics, such as joint and conditional probabilities, they argued that ultimately it is the combination of constituents *in toto* (e.g., triplets) that is the most important. It is important to note, however, that this evidence arose from spatial (not temporal) VSL.

Turk-Browne and Scholl (2009) provided evidence that representations produced from VSL are flexibly expressed, but still somewhat contingent on the order presented during familiarization. In their experiments, target triplets (e.g., ABC) were presented in a reversed order (e.g., CBA) at test, and

participants had to choose more familiar triplet set between reversed order triplets and foil triplets or target triplets, respectively. Subjects chose the reverse-ordered triplets over foil triplets at above-chance levels, but a preference for target triplets over reversed order triplet was also found. These findings indicated that 1) representations resulting from VSL generalize to some extent, with learning expressed in a flexible way, but 2) representations still maintain specific information (here, temporal order). Thus, the outcome of VSL strikes a balance between flexibility and specificity.

How flexible and general are representations derived from VSL? Extant evidence reviewed above suggests some clear limitations – seemingly, order is important to some degree, and there may be a preference for higher-order units (triplets over embedded pairs). However, the evidence is limited and, in the case of order manipulations, rests primarily on comparison at test of reordered against original orders. Further, reduced recognition of lower-order contingencies (e.g., pair components of triplets) is based solely on spatial VSL – do such differences manifest in temporal VSL? To further understand the representations arising from VSL, the present study closely examined the importance of order, pairwise association, adjacency and the potential importance of interstitial items in temporal VSL.

Similar to Fiser and Aslin (2002) and Turk-Browne and Scholl (2009), all of our experiments trained participants by presenting stimuli consecutively (i.e., temporally). In Experiments 1, we trained triplets in a temporal VSL task, and directly compared memory of triplets and each constituent pair of items from a triplet. Robust learning of embedded pairs (e.g. AB and BC in ABC) was evident, and interestingly, recognition rates of non-adjacent items (i.e. AC) also showed evidence of learning. In Experiment 2, to examine whether this learning was derived from participants' awareness of learned triplets, we not only asked participants awareness of patterns but also asked which shapes they think were associated during familiarization. The results showed no evidence of explicit awareness of specific contingencies, although some participants reported detecting regularities during familiarization. In Experiment 3, to directly test flexibility, we additionally sometimes shuffled the order of the target triplets/pairs (e.g., ACB, BAC, and CAB/ BA, CB, and CA) during the test phase. While recognition of all possible orders of target triplets and pairs were significantly higher than chance, there were no differences between canonical and corresponding randomized orderings. Additionally, AC pairs were recognized at significantly lower rates than those of three other conditions (i.e., ABC, AB, and BC). Finally, in Experiment 4, we sought to examine whether lower rates of AC recognition were due solely to distance in the triplet, or due to the fact that neither item was drawn from the temporally central "location" of the triplet (in contrast to AB and BC, which both contain the interstitial item B). By utilizing quadruplet sets to

allow for two interstitial items in each set (e.g., B and C), we compared the learning of temporally distant pairs in quadruplets (e.g., AC and BD in ABCD) to temporally distant pairs in triplets (e.g., AC in ABC). Learning for such pairs in quadruplets was found. whereas learning for AC pairs in triplets was not. We conclude that the interstitial status of B and C in quadruplets may facilitate recognition. In short, our work supports the idea of generalized and flexible representations resulting from VSL and provides evidence of how this representation is structured.

General methods

Participants

Across all experiments, 174 University of Delaware students participated for course credit or cash (Experiment 1A: N=27, Experiment 1B: N=30, Experiment 2: N=30, Experiment 3: N=27, Experiment 4: N=30 (quadruplet group)/ 30 (triplet group)). We aimed to collect 30 participants for each experiment, falling slightly short of this mark for Experiments 1A and 3 due to scheduling issues and subject drop-out.

Apparatus and materials

All experiments were run on a computer running Ubuntu Linux and attached to a 17-inch CRT monitor. All experiments were written in MATLAB using the Psychophysics Toolbox v. 3 (Brainard, 1997; Kleiner et al., 2007). As with previous research involving Western participants (Rogers, Friedman, & Vickery, 2016; Turk-Browne & Scholl, 2009; Yu & Zhao, 2015; Zhao, Al-Aidroos, & Turk-Browne, 2013), we utilized 27 symbols from the African Ndjuká syllabary as novel and unfamiliar visual stimuli. Between participants, 12 symbols were randomly selected and assigned into structured sets Fig. 1A. Stimuli were 200 pixels x 200 pixels, corresponding to approximately 5° of visual angle for participants seated approximately 57 cm from the monitor.

Procedure

Participants were given oral instructions before each experiment. Instructions were only provided for the familiarization phase of the experiment; no information was provided to participants about the test phase or the statistical relationships amongst items appearing during familiarization. After providing information about the familiarization phase of each experiment, participants were simply told to follow any additional on-screen instructions. After the participants were seated and began the experiment on their own, on-screen instructions were provided before both the familiarization phase and the test phase. The on-screen instruction of the familiarization phase was, "You will now watch a movie consisting of a series of shapes. We will ask you some questions about the movie, so pay attention to the shapes presented. Press space to start.". During the familiarization phase, participants viewed a single character at a time, always contained within a structured triplet Fig. 1B. The shapes appeared at the center of the screen, sequentially. All structured triplets were pseudo-randomized within the stream such that no immediate repetition of a triplet (e.g., ABCABC) or pair of triplets (e.g., ABCGHIABCGHI) could occur.

At the test phase, participants were asked to choose which set was more familiar between two sets (i.e., 2AFC). The instruction was, "We did not tell you this, but in the previous stage, sets of some shapes occurred together. In this final stage, we will show you two sets (or sequence for Experiment 1B). Sets (or Sequences) will consist of three shapes or two shapes. Try to guess which one the set was (or sequence) that you previously saw. If you are unsure, guess, but go for the one that feels more "familiar." Press the left arrow key if the 1st set (or sequence) is more familiar, the right arrow key if the 2nd set (or sequence) is more familiar." One set was a target set comprising a structured triplet or pair exposed during the familiarization phase (e.g., ABC or AB) and the other one was foil set comprised of recombined triplets or pairs using the first, second, and third item from different triplet or pair sets (e.g., AEI or AE). Foil sets consisted of a
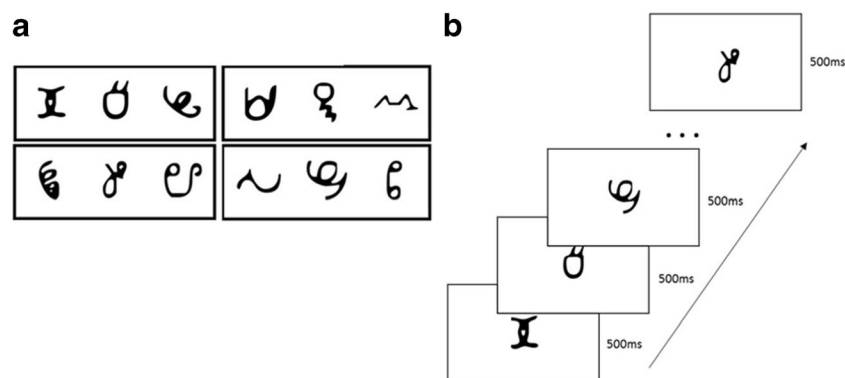

Fig. 1 Stimuli used across all experiments and familiarization phase. (A) 12 Ndjuka shapes were used. The boxes are only for demonstrative purposes. (B) General procedure of familiarization phase

stream of two or three shapes, which were generated from the shapes that used in familiarization phase but never appeared as a set. Each triplet and pair were presented against each foil, and each foil appeared as frequently as each target to prevent learning during the test phase (Turk-Browne, 2012). For the reordered target triplets or pairs (e.g., CBA or BA), reordered foils were matched (e.g., IEA or EA). A triplet (or pair) was paired with a different foil each time. The response window during the test phase was the screen with the instruction, "Press the left arrow key if the 1st set (or sequence) is more familiar, the right arrow key if the 2nd set (or sequence) is more familiar."

Experiment 1

The goal of Experiment 1 was to examine memory for pairs embedded in VSL triplets in a manner that would allow direct comparison of memory for triplets and different constituent pairs.

Methods

In Experiments 1A and 1B, a total of four triplets were predetermined, and each of them repeated 48 times during the familiarization phase. At the test phase, each target triplet and its embedded pair appeared four times (randomized order). Since there were four triplet sets and four structures (i.e., ABC, AB, BC, and AC), the total number of trials was 64. This phase was identical between Experiment 1A and Experiment 1B, with the only difference being in the presentation of stimuli during the test phase. Experiment 1A presented all of the characters of a set on screen at the same time, preceded either by "Set 1" or "Set 2". The set of triplets and pairs appeared 1500ms and 1000ms, respectively. In Experiment 1B, each structured set of characters were presented on separate screens with the words "Sequence 1" or "Sequence 2" preceding the set. Each shape was presented for 500ms Fig. 2. Experiment 1A, 2, 3, and 4 used a set presentation to test the generalizability of the representation produced by VSL. Specifically, set presentation during test might explicitly facilitate the ability to generalize, even in the absence of initial temporal information (i.e., sequential presentation in familiarization phase). We conducted Experiment 1B, however, to see whether the temporal information is necessary during test. During the test phase, one triplet (or pair) was shown on the left side of fixation, followed by another at the right side of the center. Before shapes appeared on the left (or right) side, a fixation marker was presented at the left (or right) side for 1 second to cue attention to the location. The position of the fixation marker was 250 pixels (triplet) or 350 pixels (pair) from the center (set presentation / Experiment 1A) or 200 pixels from the center (sequence presentation / Experiment 1B), which was centered on the position of the triplet/pair/sequence location.

Results

Planned comparisons against chance (50%) yielded significant learning for triplets for both Experiment 1A, $t(26) = 5.33$, $p < .001$, Cohen's $d = 1.03$ and Experiment 1B, $t(29) = 5.14$, $p < .001$, $d = 0.93$. A significant learning effect was also observed for all embedded pairs. This includes AB pairs for Experiment 1A, $t(26) = 3.69$, $p < .001$, $d = 0.71$, and Experiment 1B, $t(29) = 5.61$, $p < .001$, $d = 1.02$, BC pairs for Experiment 1A, $t(26) = 3.81$, $p < .001$, $d = 0.73$, and Experiment 1B, $t(29) = 4.90$, $p < .001$, $d = 0.9$, and AC pairs for Experiment 1A, $t(26) = 2.56$, $p = .017$, $d = 0.49$, and Experiment 1B $t(29) = 2.26$, $p = .031$, $d = 0.41$. One factor ANOVA showed a difference trending, but not significant, among the types of structure in both experiments (exp.1A: $F(3, 116) = 2.19$, $p = .09$, $\eta^2 = .06$; exp. 1B: $F(3, 116) = 2.4$, $p = .072$, $\eta^2 = .06$), but the post hoc Tukey HSD test indicated there were no differences between triplet, AB, and BC pair accuracy (all p's $> .6$), but the accuracy for AC pairs showed a non-significant trend, such that it was lower than those of three other conditions (Exp.1A: Triplet vs. AC, $p = .06$, no differences between AB vs. AC and BC vs. AC; Exp. 1B: Triplet vs. AC, $p = .1$; AB vs. AC, $p = .13$; BC vs. AC, $p = .18$) Fig. 3[1].

Discussion

Across all structure conditions, learning was above chance levels. This was particularly surprising in the case of AC pairs, since there were no interstitial items to complete an association between A and C shapes (although these were recognized less frequently than the other three conditions, at least numerically). In addition, unlike previous studies that found significant effects of structure hierarchy, such that triplets were recognized but not pairs (Fiser & Aslin, 2005; Glicksohn & Cohen, 2011), there were no significant differences between ABC, AB, and BC in our experiments. We found that triplets and their embedded pairs were learned equally well in both 1A and 1B. Therefore, our results supported the idea of pure generalized representation rather than the representation based on the hierarchical structure in VSL. Additionally, using set presentation did not yield differences between 1A and 1B, which suggests that temporal information may not be necessary in order to evoke recognition. This suggests potentially important differences between temporal and spatial VSL. In the next experiment, we examined whether this generalized

---

[1] We conducted a post-hoc ANOVA combining datasets with a between-groups factor of experiment and found no effects ($F < 1$ for main effect and all interactions).
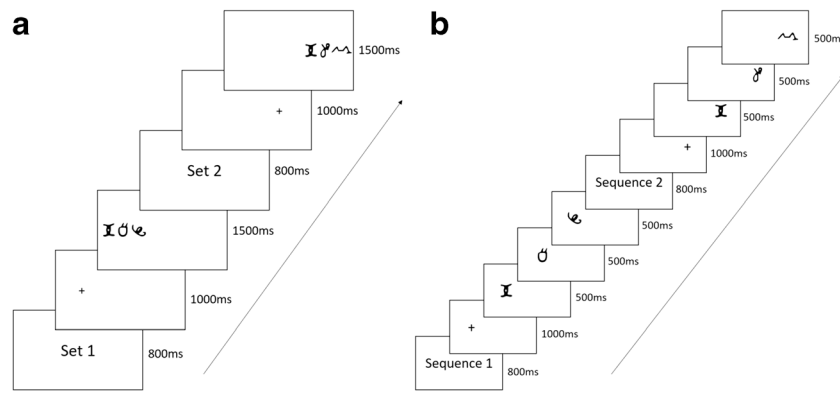
a

b

Fig. 2 General procedure of familiarity judgment task. (A) example of set presentation (used in Exp. 1A, 2, 3, and 4). (B) example of sequence presentation (used in Exp. 1B)

representation in VSL is from participants' explicit awareness of patterns of target items.

Experiment 2

Experiment 2 was undertaken to examine the role of awareness in our results, and to examine whether above-chance pair recognition may have been the result of recognizing triplets and then inferring pairs. That is, one possible reason for pair recognition in Experiment 1 is that participants may have recognized a triplet against a foil (e.g., ABC vs. AEI), then subsequently chosen any constituent combination of A, B, and/or C when they appeared together. To address these issues we conducted Experiment 2, in which we asked participants questions about general awareness (of the purpose of the experiment, and about patterns present during familiarization), and further probed their explicit awareness of contingencies by directly asking them to identify associates of each shape, following the familiarization stage but preceding the familiarity test phase. Finally, we removed triplets from the familiarity test such that pair recognition could not be based on explicit triplet recognition from the test phase, alone.
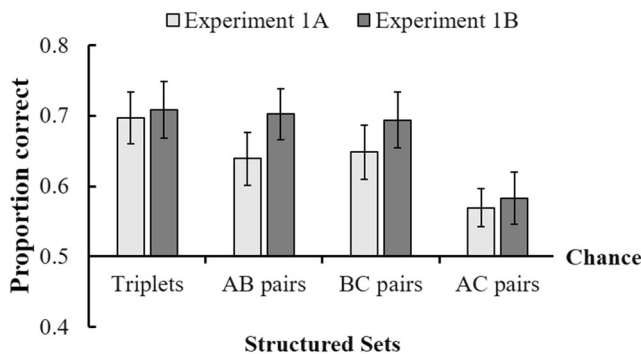

Fig. 3 Accuracy at choosing target triplets and pairs (AB, BC and AC) over foil triplets and pairs in Experiment 1. Only the presentation type was different between Experiment 1A and 1B: set presentation (exp.1A) and sequence presentation (exp.1B). In this and all other figures, error bars represent standard error of the mean

Methods

Experiment 2 was identical to Experiment 1A, except as follows. After the familiarization phase and before moving on to the test phase, participants were asked their awareness of any patterns of shapes during the familiarization phase. They were first asked these four questions: 1) What do you think the experiment was about? 2) Have you encountered an experiment like this before? 3) Did you notice any repeating patterns (YES or NO)? If yes, what were the patterns? What is your confidence that there were patterns in the stream? (5-point likert scale), 4) If you had to guess between these options, would you guess that shapes were presented in PAIRS, TRIPLETS, QUADRUPLETS, or higher combinations of shapes? What is your confidence of your answer above? (5-point likert scale). After they answered for these questions, participants were instructed to perform a matching questionnaire (Kim, Seitz, Feenstra, & Shams, 2009). The instruction of this task was to "choose which of eleven shapes are related to the example shape based on what you saw during the first phase of the experiment." In this phase, one shape at a time (randomized order) was presented at the top of the screen and the rest of the 11 shapes were presented at the bottom of the screen, labeled from 1 to 11. Since triplets were presented during the familiarization, there were two correct answers among eleven shapes for each example shape. Following this phase was the same test phase that was used in Experiment 1A (i.e. 2AFC). However, we only measured recognition of canonically-ordered embedded pairs (e.g., AB, BC, and AC) to prevent participants from inferring the embedded pairs from triplets during the test phase. These embedded pairs appeared four times (48 trials, total).

Results

We again replicated our results such that planned comparisons against chance (50%) yielded significant learning for all embedded pairs, AB: t(29) = 5.91, p < .001, Cohen's *d* = 1.08,

BC: t(29) = 4.91, p < .001, $d = 0.9$, and AC: t(29) = 2.12, p = .04, $d = 0.38$ Fig. 4. One-factor ANOVA showed a significant difference among the structures, F(2, 87) = 3.15, p = .048, η_p^2 = .07). The post hoc Tukey HSD test indicated there were no differences between AB and BC pairs and BC and AC pairs in their proportion correct (all p's > .2), but the proportion correct for AC pairs were significantly different from AB pairs, t(29) = 2.42, p = .045, $d = 0.256$.

On the four questionnaires, 75% of participants (20 out of 30 people) correctly guessed that the experiment involved memories of shapes, and some mentioned memory for patterns of shapes (question #1), and 17% of participants reported that they have encountered experiment like this before (question #2). The critical questions were question #3 and #4, and surprisingly, 100% people reported that they noticed repeating patterns. However, only 47% of participants (14 out 30 participants) correctly reported that they saw triplets during the familiarization.

We then calculated d prime based on participants' responses on the shape-matching questionnaire. We divided responses into four types: hits (items which participants correctly associated to the example shape), false alarms (items which they incorrectly associated to the example shape), miss (items which they did not report even though they should have), and correct rejections (items which they correctly did not report as associated with the shape). The sensitivity index, d prime, was derived from these statistics, to examine whether participants correctly discriminated target items from non-target items (Brophy, 1986). We observed a negative d prime for most of the participants (27 out of 30), meaning that overall, the false alarm rate greater than the hit rate (M = -1.25, SD = 2.57). We divided participants who reported: "there was a triplet" from others ("there was pair/ quadruplet/ higher combination") on question #3 and found that both groups exhibited a negative d' (triplet: M = -0.76, SD = 3.66; other than triplet: M = -1.69, SD = 0.82), and there was no difference between these two groups in terms of d' (t < 1). Finally, comparing recognition rates between these groups using ANOVA resulted in no main effect or interaction involving group (all F < 1). Bayesian Repeated Measures ANOVA (using a group as a factor) showed evidence favored the null hypothesis for effects involving groups
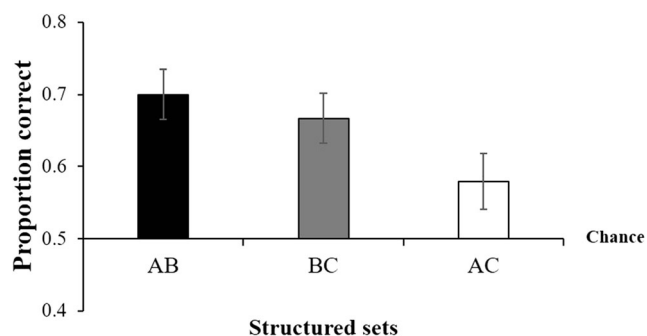
by at least a 2:1 margin (group main effect, BF_{01} = 2.61; interaction, BF_{01} = 5.73), indicating evidence against the possibility of different accuracy between groups (Love et al., 2015). We also examined the d' of participants who reported, "there was a pair." 23% of participants (7 out 30 participants) reported that they saw pairs during the familiarization. Again, all of them exhibited a negative d' (pair: M = -1.69, SD = 0.95), and when comparing recognition rates among participants who reported "pair" vs. "triplet" and "pair" vs. others ("there was quadruplet/higher combination") using ANOVA, the results showed no main effect or interaction involving group (all F < 1). Thus, those with partial knowledge of the true structure exposed during familiarization also showed no learning advantage.

Discussion

The present experiment provided evidence that memory of pairs was likely not derived from explicit memory of learned structures. While all participants reported awareness that there were patterns in sequences, and half of them correctly reported that the sequence was composed of triplets, they were nevertheless poor at explicitly recognizing which particular shapes were associated. In addition, by only using embedded pairs at the test phase, we avoided the possibility of inference of pairs from triplets during the test phase. Participants achieved above-chance performance for all types of constituent pairs, even AC pairs.

In Experiments 1 and 2, we found robust recognition of pairwise associations among the components of a triplet. We next considered how various permutations of characters within a structured set may be treated after VSL, in order to further test the flexibility of units resulted from VSL.

Experiment 3

Evidence from Experiment 1 and 2 suggested VSL produces a generalizable representation in which even A and C items are judged as familiar. We next sought to test how flexible these representations are to different orderings, by testing permutations of structured items.

Methods

Experiment 3 was identical to Experiment 1A, with the exception that every possible triplet and target pair permutation was presented during the test phase. Each target triplet and all possible permutations (i.e., six possibilities for each triplet: ABC, ACB, BAC, BCA, CAB, and CBA), its embedded pairs and their reversed order (i.e., total six variations, AB, BA, BC, CB, AC, and CA) appeared four times. Therefore, the total trials of the test phase were 192.

Fig. 4 Proportion correct of choosing correct embedded pairs (AB, BC and AC) over foil pairs in Experiment 2

Results

Learning was evident with all structured set types being recognized over foils above chance[2]. Paired Samples t-test revealed no differences between the canonical order of triplets (e.g., ABC) (M= 0.68, SD=0.04) and the mean of the scrambled order of triplets (e.g., ACB, BAC, BCA, CAB, and CBA) (M=0.64, SD=0.04); t(26) = 0.59, p = .56. Bayesian Paired Samples t-test also suggested that evidence favored the null hypothesis (BF_{01} = 4.34). In addition, each canonical order pair and its reversed order pair (i.e., AB vs BA, BC vs. CB, and AC vs. CA) did not yield any statistical differences, either (all t < 1). Bayesian Paired Samples t-test again showed evidence favored the null hypothesis for all comparisons (AB vs. BA: BF_{01} = 4.73; BC vs. CB: BF_{01} = 1.67; AC vs. CA: BF_{01} = 4.9). An ANOVA showed no differences among the structures (F < 1), and the post hoc Tukey HSD test also indicated there were no differences between structures (all p's > .9) Fig. 5. We examined a strength of evidence of structures by applying Bayesian ANOVA (using a type of structures as a factor), and the results showed evidence strongly favored the null hypothesis (structures main effect, BF_{01} = 233.13), indicating strong evidence against the possibility of different recognition rate among structures.

Discussion

Our results demonstrate representations resulting from VSL to be highly flexible and generalizable, as evidenced by every permutation being judged as more familiar than foils at above-chance levels. Again, we found A and C pairs are learned as a pair, although the magnitude of this learning is numerically less than other structures (i.e., triplets, AB and BC pairs).

It remains possible that some of the above-chance recognition rates in Experiment 3 may be due to generalization taking place during the test phase. We view this as unlikely, since Experiment 2 showed no apparent reduction in pair recognition rates in the absence of triplet tests, and because if this were taking place routinely during the test phase, we would still expect lower familiarity for reversed pairs / shuffled triplets, as presumably one would not recognize randomly-ordered combinations (BCA, CBA, etc.) until tested with the original order (ABC).

---

[2] This included base (ABC) triplets, t(26) = 4.00, p < .001, Cohen's d = 0.77, ACB triplets, t(26) = 3.55, p = .002, d = 0.68, BAC triplets, t(26) = 4.35, p < .001, d = 0.84, BCA triplets, t(26) = 2.86, p = .008, d = 0.55, CAB triplets, t(26) = 2.74, p = .011, d = 0.53, CBA triplets, t(26) = 4.25, p < .001, d = 0.82. Additionally, all pairs reached significance with AB pairs, t(26) = 4.25, p < .001, d = 0.82, BA pairs, t(26) = 5.19, p < .001, d = 1, BC pairs, t(26) = 3.90, p < .001, d = 0.75, CB pairs, t(26) = 3.04, p = .005, d = 0.59, AC pairs, t(26) = 2.66, p = .013, d = 0.51, and CA pairs t(26) = 2.43, p = .022, d = 0.47.

Experiment 4

In Experiment 1A, 1B, 2, and 3, we found numerically worse recognition performance for AC compared to other structures, although AC recognition was inconsistently worse statistically. We speculated that 1) distance within structured items may impact the degree of recognition, and/or 2) the presence of an interstitial item aids recognition. The latter possibility is quite plausible as we found learning effects even with completely randomized triplet order, but the AC and CA pairs continued to show relatively reduced recognition rates. Experiment 4 was designed to separate effects of distance from those resulting from the presence of an interstitial item, by comparing pair learning of pairs separated by a single item across triplets and quadruplets.

Methods

In Experiment 4, participants were pseudorandomly assigned to one of two groups, a "quadruplets" group (N=30) that were exposed to four quadruplets (composed of 16 total shapes) and a "triplets" group (N=30) exposed to five triplets (composed of 15 total shapes). Due to time constraints we cut the total number of trials in half during the familiarization phase compared to Experiments 1-3. We also speculated that this reduced training regimen might enhance differences between AC and AB/BC recognition. In this phase, four quadruplets or five triplets were presented and each of them repeated 24 times in a pseudorandom order. The test phase was identical to Experiment 1A, but we only measured recognition of canonically-ordered embedded pairs (e.g., AB, BC, CD, AC, BD, and AD for quadruplets; AB, BC, and AC for triplets). These pairs appeared four times or five times, respectively (total trials for quadruplet: 96; triplet: 75).

Results

In the group exposed to four structured sets of four items, learning was evident with all but one condition above chance, including AB pairs t(29) = 4.07, p < .001, Cohen's d = 0.74, BC pairs, t(29) = 3.14, p = .004, d = 0.57, CD pairs, t(29) = 3.01, p = .005, d = 0.55, AC pairs, t(29) = 2.70, p =.012, d = 0.49, and BD pairs, t(29) = 2.37, p = .024, d = 0.43. Only the structured pair that possessed no interstitial items, AD pairs, failed to reach significance t(29) = 0.92, p = .366, d = 0.17.

Results from the group with five structured sets of three items also highlighted the importance of interstitial items with significant learning observed for AB pairs, t(29) = 3.02, p = .005, d = 0.55, and BC pairs, t(29) = 3.86, p < .001, d = 0.7, but not AC pairs, t(29) = -0.309, p = .759, d = -0.056 Fig. 6.

Critically, we next tested whether items with equivalent spacing were correctly judged as more familiar if they
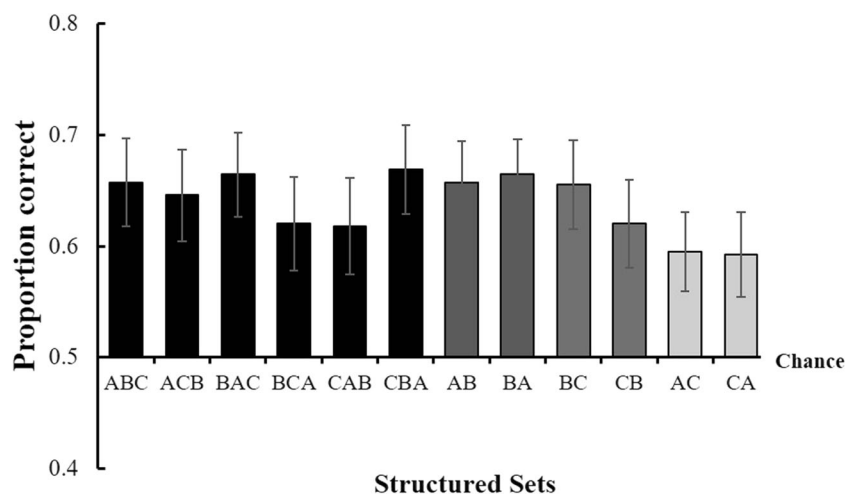
Fig. 5 Accuracy in familiarity test for target triplets and pairs (AB, BC and AC) and randomly ordered triplets (ACB, BAC, BCA, CAB, and CBA) and pairs (BA, CB, and CA) over foil triplets and pairs when the shapes presented as a set during the familiarity judgment task

contained an interstitial item (i.e., AC/BD in the quadruplet group, which contain interstitial items C and B, respectively) than those that did not (i.e., AC in the triplet group). We conducted planned comparisons, and the results showed that AC and BD in ABCD (i.e., Quadruplet condition) were both associated with significantly higher recognition rates than AC in ABC (i.e., Triplet condition) (AC vs. AC, $t(58) = -2.38$, $p = .02$, $d = -0.61$; BD vs. AC, $t(58) = -2.05$, $p = .045$, $d = -0.53$).

Discussion

Our results support the idea that interstitial items are essential in VSL, such that only the quadruplet condition (e.g., ABCD) produced significant learning for AC and BD pairs, where B and C are now both interstitial items, whereas AC learning was not observed in the triplet group. In other words, the generalized representation produced by VSL may rely on an association between items. The remote pairs in triplet groups (e.g., AC in ABC) mirrored the quadruplet AD condition where an absence of interstitial items is associated with unsuccessful recognition.

We did not observe above-chance recognition of AC pairs from triplets as we did in our previous experiments. In considering why this is the case, it is important to mention that this experiment and previous experiments differ in that 1) the number of triplet sets was greater in Experiment 4 (Experiments 1-3: 4 sets; Experiment 4: 5 sets), and 2) the number of exposures to the stimuli was lower in Experiment 4 (Experiment 1-3: 48 times total before test, Experiment 4: 24 times). It is possible that additional familiarization time is required for AC recognition to manifest. Therefore, we conducted a supporting experiment, such that 5 sets of triplets were shown 48 times and see whether we can find the AC learning with increased learning time (N=21). The results

again showed the AC learning, supporting the idea of generalized representation in VSL[3]. There was no significant difference between AC and AB or BC (both t <1). A limitation of this study is that it is inherently difficult to compare triplet with quadruplet learning, since there is no way to simultaneously equate the total number of constituent items used as well as the number of structured units (quadruplets or triplets).

General discussion

Our study examined the extent to which VSL supports the recognition of pairs of items that are separated by other elements during learning ("remote pairs"), and the extent to which recognition is tolerant to different orderings of items compared to experiences during learning. The experiments reported here not only demonstrate a robust and replicable learning of remote pairs of items (Experiments 1) but also provided evidence that recognition abstracted over initially presented orderings (Experiment 3). These findings were not based on the participants' explicit awareness of contingencies, nor were they induced at test due to recognition of triplets (Experiment 2). Moreover, we highlighted the fact that interstitial items may play a special role in representations resulting from VSL (Experiment 4), as indexed by superior recognition of remote pairs that include at least one such item compared to those that only contain boundary items. These results supported the idea that representation produced by VSL are flexible and generalizable.

Fiser and Aslin (2002) and Turk-Browne and Scholl (2009) both observed evidence that temporal order information in

---

[3] Planned comparisons against chance (50%) yielded significant learning for all pairs. This includes AB pairs, $t(20) = 2.82$, $p = .01$, $d = 0.62$, BC pairs, $t(20) = 2.4$, $p = .03$, $d = 0.52$, and AC pairs, $t(20) = 2.3$, $p = .03$, $d = 0.5$
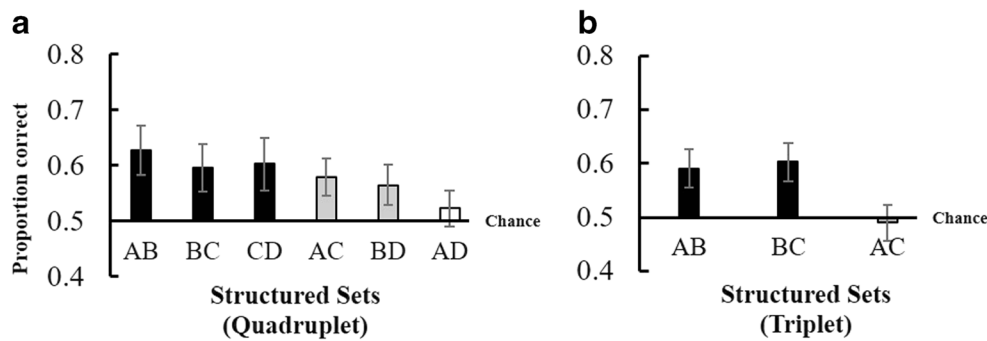
Fig. 6 Accuracy in choosing embedded pairs over foil pairs in Experiment 4. (A) embedded pairs from quadruplet sets (e.g., ABCD). (B) embedded pairs from triplet sets (e.g., ABC)

memory manifests when that information is all that discriminates a target from a non-target. Fiser and Aslin (2002) argued that VSL of structures composed of more than two items is based upon first-order temporal statistics, such that when people were asked to choose the most familiar between BCA and ABC, there was no differences in performance because both of structures started with a canonically-ordered pair (i.e., BC and AB). In contrast, discriminations comparing CAB and ABC resulted in a preference for ABC, apparently because the first pair of CAB did not contain first-order statistical contingency. Similarly, Turk-Browne and Scholl (2009) provided evidence that temporal order affected the expression of learning (i.e., the recognition rate of base triplet was significantly above chance when compare with its reversed order, ABC > CBA). However, it did not affect learning when that information was not needed while testing (i.e., CBA > foil). We have replicated the findings of Turk-Browne and Scholl (2009), in the sense that we found that scrambled orders where recognized when pitted against foils and extended their findings by scrutinizing how broadly generalization can be applied and what parts of the representation are critical to the integrity of recognition. However, our findings do not support the claim that VSL representations are based on first-order temporal statistics, because there was no advantage at recognition for canonical orders compared with other orderings (e.g., ABC, BCA = ACB, BAC, BCA, CAB, CBA). Additionally, we found above-chance recognition for all variations of pairs and triplets, which is not consistent with findings of Fiser and Aslin (2005) and Glicksohn and Cohen (2011).

We propose two possible explanations for why we were able to find evidence of generalizability in cases where previous studies have not. First, all of our experiments trained participants by presenting stimuli temporally, while Fiser and Aslin (2005) and Glicksohn and Cohen (2011) presented stimuli concurrently (i.e., spatially, at the same time). This may have impacted how VSL led to the creation of group representations. In other words, spatial VSL may operate by different rules than temporal VSL. For example, learning temporal sequences more likely to focus on the link (or association)

between items (e.g., Spiegel & McLaren, 2006); whereas learning spatial information focuses on extracting objects as a chunk (e.g., Fiser & Aslin, 2005). Therefore, it is possible that the generalized representation of VSL may only be shown in the temporal domain, not spatial, because when a stream of items was sequentially presented during learning in the temporal domain (not presented at the same time), people may process a link between the items, not the exact group of shapes. The nature of this associative learning may be distinct from the kind of group learning that guides spatial VSL. Future work will be needed to examine whether spatial and temporal VSL differ in this regard.

Secondly, the experimental design in the familiarization phase of Fiser and Aslin (2002) might emphasize different aspects of temporal information compared with our experimental design. While ours was a simple presentation of a stream of shapes presented one at a time, Fiser and Aslin used items that moved back and forth behind an occluder, such that shape identities changed when they reappeared on the other side of the occluder. It is possible that this kind of evolution evoked different mechanisms of learning, or emphasized unique aspects of temporal contingency, compared with our design. Lastly, amount of exposure may be critical, with longer and/or greater number of exposures leading to stronger and more generic representations. Supporting this possibility, we failed to observe significant recognition rates for remote pairs in triplets only once out of five experiments, specifically when the number of exposures was limited. Unlike Experiment 1, where stimuli were presented 48 times total before test, Experiment 4 only presented triplets 24 times, leading to a weakened representation. It seems clear that, while interstitial items are important for the recallability of VSL representations, their absence does not prevent learning and expression, entirely. In other words, Fiser and Aslin (2002, 2005) and Glickson and Cohen (2011) may not have provided sufficient training for participants to build representations that were immune to order effects, just as observed in Experiment 4. Given that there were some differences between previous studies and our study, our

findings suggest a new perspective on the generalization of VSL representations. Consistent with what Turk-Browne and Scholl (2009), robust recognition of not only reversed order (i.e., CBA), but completely randomized order of triplets and pairs does not support the idea of representation determined by first-order temporal statistics.

To address participants' explicit knowledge about structures, we measured participants' degree of awareness more directly. The results showed no evidence of explicit awareness of contingencies (i.e., which shape was associated with which other shape), and some evidence that participants were generally aware of some patterns during familiarization. Due to this mixed evidence of awareness, it is hard to make strong conclusions regarding whether the learning effects observed here are (in part or in whole) "implicit" or "explicit." The issue of whether VSL is implicit or explicit is complex; for example, VSL may or may not depend on awareness of regularity during familiarization, but even if it does, the resulting memory may often be inaccessible to conscious report. These broad issues are beyond the scope of this paper, but deserve further study.

We additionally found evidence that suggests that interstitial items are privileged, such that pairs containing interstitial items were better recognized than those that do not; recognition of AC and BD in ABCD quadruplets was better than that of AC pairs from ABC triplets. We assume that this pattern of results was due to the fact that B and C items in ABCD are interstitial items. Turk-Browne and Scholl (2009) suggested generalized representations in VSL can be possible because there might be an associative component involved in the process of learning regularities. By showing that regularities learned in time can be expressed spatially and vice versa, they argued that representations formed from VSL are learned associations. In line with this claim, our findings in Experiment 4 support the idea of an important role of associative components in the process of learning regularities, such that even with items that are non-adjacent to each other, as long as one of those items was interstitial item, it resulted in robust learning effect (e.g., AC and BD in ABCD). In other words, linking items which are essential for constructing associations during familiarization facilitated recognition.

Extending the findings of generalized representations in Turk-Browne and Scholl (2009), our study provided evidence that these representations may be much more generalized (e.g., remote pair learning) and flexible (e.g., completely randomized order) in temporal VSL than previously appreciated. These findings add important knowledge about the flexibility and generalizability of VSL, to a limited literature that specifically examines such questions. Further, we find that interstitial items are critical to VSL, which suggests a novel and key role for such items in binding together elements that are separated by interstitial items. Future research should examine the underlying mechanisms that emphasize interstitial items, and the downstream consequences (e.g., to attention) of their presence.

References

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brophy, A. L. (1986). Alternatives to a table of criterion values in signal detection theory. Behavior Research Methods, Instruments, & Computers, 18(3), 285–286.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M. (2003). Scene perception and memory. In *Psychology of Learning and Motivation* (Vol. 42, pp. 79–108). Elsevier.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

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*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology. General*, *134*(4), 521–37. https://doi.org/10.1037/0096-3445.134.4.521

Glicksohn, A., & Cohen, A. (2011). The role of Gestalt grouping principles in visual statistical learning. *Attention, Perception & Psychophysics*, *73*(3), 708–13. https://doi.org/10.3758/s13414-010-0084-4

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., & Wagenmakers, E. J. (2015). JASP (Version 0.7)[Computer software]. Amsterdam, The Netherlands: JASP Project. Retrieved from https://jasp-stats.org.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual Statistical Learning Based on the Perceptual and Semantic Information of Objects. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *39*(1), 196–207. https://doi.org/10.1037/a0028645

Rogers, L. L., Friedman, K. G., & Vickery, T. J. (2016). No Apparent Influence of Reward upon Visual Statistical Learning. *Frontiers in Psychology*, *7*, 1687.

Spiegel, R., & McLaren, I. P. L. (2006). Associative sequence learning in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 156.

Turk-Browne, N. B. (2012). Statistical learning and its consequences. In *The influence of attention, learning, and motivation on visual search* (pp. 117–146). Springer.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–64. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 195–202. https://doi.org/10.1037/0096-1523.35.1.195

Yu, R. Q., & Zhao, J. (2015). The persistence of the attentional bias to regularities in a changing environment. *Attention, Perception, & Psychophysics*, *77*, 2217–2228. https://doi.org/10.3758/s13414-015-0930-5

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention Is Spontaneously Biased Toward Regularities. *Psychological Science*, *24*(5), 667–677.