

Multidimensional Visual Statistical Learning

Nicholas B. Turk-Browne, Phillip J. Isola, Brian J. Scholl, and Teresa A. Treat
Yale University

Recent studies of visual statistical learning (VSL) have demonstrated that statistical regularities in sequences of visual stimuli can be automatically extracted, even without intent or awareness. Despite much work on this topic, however, several fundamental questions remain about the nature of VSL. In particular, previous experiments have not explored the underlying units over which VSL operates. In a sequence of colored shapes, for example, does VSL operate over each feature dimension independently, or over multidimensional objects in which color and shape are bound together? The studies reported here demonstrate that VSL can be both object-based and feature-based, in systematic ways based on how different feature dimensions covary. For example, when each shape covaried perfectly with a particular color, VSL was object-based: Observers expressed robust VSL for colored-shape sub-sequences at test but failed when the test items consisted of monochromatic shapes or color patches. When shape and color pairs were partially decoupled during learning, however, VSL operated over features: Observers expressed robust VSL when the feature dimensions were tested separately. These results suggest that VSL is object-based, but that sensitivity to feature correlations in multidimensional sequences (possibly another form of VSL) may in turn help define what counts as an object.

Keywords: statistical learning, feature binding, objects, feature dimensions

Visual perception is remarkable in at least two ways. First, it serves to make us aware of a highly coherent and structured world, despite noisy fragmented input. Second, it provides such experiences without any hint of the underlying computational complexity involved in their construction. In fact, recent work has demonstrated that perception is supported by surprisingly subtle types of visual associative learning. One type of learning—visual statistical learning—may be particularly important in this context, since it can occur automatically and without any intent or even awareness.

Visual Statistical Learning

The study of implicit learning has a long history in psychology (see Stadler & Frensch, 1998), going back to early studies of learning in natural languages (e.g., Harris, 1955), artificial grammars (e.g., Reber, 1967), and manual sequences of responses (e.g., Nissen & Bullemer, 1987). In its modern incarnation, the study of statistical learning began with the demonstration that young infants are able to find “word” boundaries in acoustically unsegmented syllable sequences by using only the differential transitional prob-

abilities between syllables (Saffran, Aslin, & Newport, 1996). The precise relationship between such forms of statistical learning and the implicit learning literature more generally is under active consideration (Perruchet & Pacton, 2006).

The auditory statistical learning design was later adapted for use with visual stimuli in studies with adult observers by Fiser and Aslin (2002a; see also Olson & Chun, 2001). Observers viewed an animation in which a single object moved horizontally across the screen, continuously cycling back and forth behind a central occluder and changing its shape each time it passed behind the occluder (see Figure 1a). Observers watched this animation for only a few minutes, with no specific task. The sequence of shapes, though apparently random, actually consisted of temporal “triplets” in which the same three shapes always appeared in the same order (e.g., A-B-C-G-H-I-D-E-F-A-B-C . . .). Critically, only this statistical regularity demarcated the triplets, since the intershape delay was always the same.

After this passive exposure, observers completed a surprise two-interval-forced-choice familiarity task that pitted triplets (e.g., ABC) against foil sequences of three shapes with a joint probability of zero (e.g., AED). Observers correctly identified the triplets as more familiar than the foil sequences 95% of the time, indicating robust statistical learning of visual temporal sequences. While the use of deterministic triplets and familiarity judgments may support explicit recognition of the regularities in some cases, visual statistical learning (VSL) can also occur during an orthogonal task (such as repetition detection), without any reported awareness of the statistical structure (Turk-Browne, Jungé, & Scholl, 2005), and can facilitate online performance (Hunter & Aslin, 2001; Olson & Chun, 2001; Turk-Browne et al., 2005; Turk-Browne & Scholl, 2006).

Other recent work with VSL has demonstrated that such learning operates over spatial as well as temporal regularities (e.g.,

Nicholas B. Turk-Browne, Phillip J. Isola, Brian J. Scholl, and Teresa A. Treat, Department of Psychology, Yale University.

Nicholas B. Turk-Browne was supported by a foreign Natural Sciences and Engineering Research Council of Canada postgraduate scholarship, and Brian J. Scholl was supported by National Science Foundation Grant BCS-0132444. For helpful conversation and comments on drafts of the article, we thank Tim Brady, Marvin Chun, Rochel Gelman, Justin Jungé, Stephan Lewandowsky, Christian Luhmann, and Pierre Perruchet. We thank Dick Aslin for providing the shape stimuli.

Correspondence concerning this article should be addressed to Nicholas B. Turk-Browne or Brian J. Scholl, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. E-mail: nicholas.turk-browne@yale.edu or brian.scholl@yale.edu

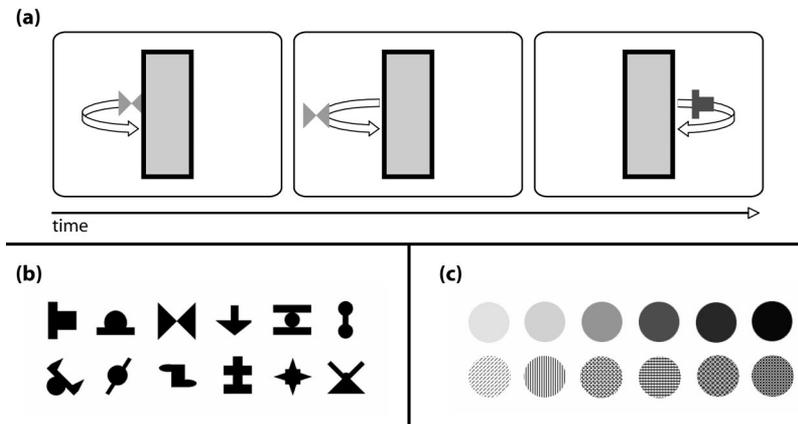


Figure 1. Stimuli and trial sequence. (a) Depiction of the display used in the present experiments (and in Fiser & Aslin, 2002a). A single object oscillates back and forth across the display, changing into a new object each time it passes behind the stationary central occluder. (b) The 12 nonsense shapes used in our study, from Fiser and Aslin (2001). (c) The 12 nonblack colors used in our study (depicted here with different patterns and gray levels).

Baker, Olson, & Behrmann, 2004; Chun & Jiang, 1998; Fiser & Aslin, 2001); at multiple spatial scales (Fiser & Aslin, 2005); across multiple modalities (Conway & Christiansen, 2005) and despite interleaved noise (Jungé, Turk-Browne, & Scholl, 2005; Turk-Browne et al., 2005); and in young infants in addition to adults (Fiser & Aslin, 2002b; Kirkham, Slemmer, & Johnson, 2002). Moreover, other recent studies of temporal VSL have begun to elucidate some of the underlying processes that help make VSL possible, involving selective attention (Turk-Browne et al., 2005), association (Turk-Browne & Scholl, 2006), computations of persisting object representations (Fiser, Scholl, & Aslin, 2007), and anticipation (Turk-Browne, Johnson, Chun, & Scholl, 2007).

The Units of VSL: Features or Objects?

Despite this considerable body of work on VSL, several fundamental questions remain about its nature. In particular, previous experiments have not determined the underlying units over which VSL operates. In a sequence of colored shapes, for example, does VSL operate over each feature dimension independently, or over multidimensional objects in which color and shape are intrinsically bound together? Our study modifies two aspects of previous designs, which rendered them unable to answer this question. First, most previous studies used either monochromatic shapes (so that there was no statistical information about color; e.g., Fiser & Aslin, 2002a) or sequences in which each shape had a single unique color (so that the statistics of color were also perfectly correlated with those of shape; e.g., Fiser & Aslin, 2002b; Kirkham et al., 2002). Second, in all but one previous VSL experiment, the surface features present during familiarization were always present during test.¹

Exploring the nature of the underlying units of VSL seems important for several reasons: First, the existence of multiple feature dimensions is an inescapable facet of real-world visual experience, and thus these studies may help to reveal how VSL operates in a more ecologically valid situation. Second, exploring the necessity of certain features during familiarization and test may

have important implications for how readily learning in one situation will transfer to another (see also Turk-Browne & Scholl, 2006). Finally, and most generally, one of the most critical steps in understanding any cognitive or perceptual process is to determine the underlying currency over which that process operates. In the following experiments, we explore how VSL operates over individual features and bound objects.

General Method

Observers

Seventy-two naïve subjects (16 in Experiments 1, 2, 3, and 4a; 8 in Experiment 4b), all with normal or corrected-to-normal acuity and color vision, participated for course credit or monetary compensation.

Apparatus and Stimuli

Stimuli were presented with custom software written with the VisionShell graphics libraries (Comtois, 2004) on Apple desktop computers. Observers sat approximately 46 cm from the monitor without head restraint. Displays were constructed from a set of 12 nonsense shapes used previously in studies of VSL (Figure 1b; e.g., Fiser & Aslin, 2001). Each shape appeared in one or more of 13 colors (including black; see Figure 1c), and subtended roughly 3.3° . Each sequence began with the object located in the center of

¹ The one exception is the study of Turk-Browne et al. (2005), who observed expression of learning at test for monochromatic shapes that had been colored during familiarization. However, the color information during familiarization in this study was used only to manipulate attention, and was not diagnostic; in fact, each shape throughout the study appeared in one of only two colors, and only one color was attended. In addition, one auditory statistical learning study has employed stimuli that varied between familiarization and test: Thiessen and Saffran (2003) changed the prosody of spoken syllables after familiarization and still observed auditory temporal statistical learning in 9-month-old infants.

the display, where it immediately began moving to either the left or right at 11.48°/s. When the center of the object was 12.76° from the display border, the object immediately reversed direction and moved toward the same point on the opposite side of the screen. The object then continued back toward the center of the display, at which time the entire movement could be repeated. A stationary gray occluder, subtending $9.51^\circ \times 5.68^\circ$, was always present in the center of the display. The object was (at least partially) visible for 1 s between each moment of complete occlusion (see Figure 1a).

Procedure

Familiarization phase. Observers watched a 5-minute sequence of 288 colored shapes, presented as described above. Although the sequence of colors varied across experiments, the sequence of shapes was always identically structured. Each shape was assigned to a unique position in one of four triplets—sequences of three shapes that always appeared in the same order (e.g., ABC, GHI, etc.). Each third of the full shape sequence was generated by randomly interleaving 8 repetitions of each triplet with two constraints (for a total of 24 repetitions/triplet): (a) no triplet could be repeated sequentially (e.g., ABCABC), and (b) no pair of triplets could be immediately repeated (e.g., ABCGHIABCGHI). As a result, the joint probability of a triplet (e.g., ABC) was 0.083, while the joint probability of a sequence of three shapes spanning triplets (e.g., CGH) was 0.027.²

Test phase. After the familiarization phase, observers completed 32 two-interval forced-choice test trials, judging the relative familiarity of (a) triplets versus (b) foil sequences of three familiar shapes that had never appeared sequentially during familiarization.³ The triplet and foil sequences were presented on each trial in the same disoccluding–occluding manner as during familiarization, separated by a 1-second pause, and were randomly chosen with equal likelihood; the presentation order of the alternatives was randomized. Observers pressed a key if the first sequence seemed more familiar and a different key if the second sequence seemed more familiar.⁴

Experiment 1: Testing Bound Objects

While most previous studies have employed monochromatic shapes, this preliminary experiment verified that our method could give rise to VSL even when each shape had its own unique color during both familiarization and test, as a critical baseline for our later experiments.

Method

Each shape was assigned a single unique nonblack color, in which it always appeared throughout the experiment (Figure 2a).

Results and Discussion

In this and all future experiments, our measure of VSL was the percentage of test trials in which the triplet was chosen over the foil (where chance = 50%). Observers showed robust statistical learning (80%; Figure 3a) that differed significantly from chance, $t(15) = 5.730, p < .001, d = 1.432$. This effect serves as a baseline for later experiments.

Experiment 2: Testing Separated Features

Did VSL in Experiment 1 operate over bound object representations? If so, then the presence of both features should be necessary for learning to be expressed at test, such that testing either monochromatic shapes or colored circles would break the binding (i.e., the associations between dimensions) and should attenuate the expression of VSL.

Method

The familiarization phase of this experiment was identical to that of Experiment 1. The test phase consisted of two separate blocks: a shape test and a color test (with order counterbalanced across observers). The shape test was identical to the test in Experiment 1, except that all shapes were presented in black (such that only shape information was preserved; see Figure 2b). Conversely, the color test was identical to the test in Experiment 1, except that all colors were presented as circles (such that only color information was preserved; see Figure 2b).

Results and Discussion

As depicted in Figure 3b, observers expressed weak VSL in the shape test (57%), $t(15) = 2.159, p = .047, d = 0.540$, and these results were significantly weaker than those in Experiment 1, $t(30) = 3.699, p < .001, d = 1.308$. Observers expressed no significant VSL in the color test (55%), $t(15) = 1.412, p = .178, d = 0.353$, and these results were again significantly weaker than those in Experiment 1, $t(30) = 4.033, p < .001, d = 1.426$. There was no difference between the performance for shape and color ($t < 1$). Critically, the familiarization phase in this experiment was identical to that in Experiment 1, suggesting that the weaker results here reflect only the expression of learning rather than the learning itself. These results suggest that performance in Experiment 1 was not driven by learning of sequences of features in either the shape or color dimension alone. Given that both features were necessary for VSL to be expressed, learning must have operated over bound shape–color pairs.

An alternative explanation could be that individual observers only ever attended to either shape or color during learning. Thus,

² For derivations of these probabilities, see Turk-Browne et al., 2005.

³ The foils (e.g., AEI) contained three shapes that were equally familiar as the shapes in the triplets, but they had never been presented in succession during familiarization. Thus, the two test sequences in each trial could be discriminated only based on statistical learning of the joint probabilities between shapes during familiarization: Observers could respond accurately only by noting relationships between individual shapes, in particular by encoding their nonzero joint probabilities. Note that while this is perhaps the most direct test of statistical learning, other studies have also tested for the learning of even subtler statistics, such as relative joint probability and conditional probability.

⁴ Following the test phase, observers were carefully debriefed to assess their explicit awareness of the triplet structure from familiarization. Note that such awareness would not necessarily indicate explicit or strategic learning per se; observers may instead become aware of the structure after implicit processes have run their course. Regardless, only 2 of 72 observers reported becoming fully aware of the triplet structure during familiarization. The majority of subjects fell into one of two groups, either reporting no awareness of any regular sequences or reporting that they noticed one or more pairs of objects that consistently appeared together.

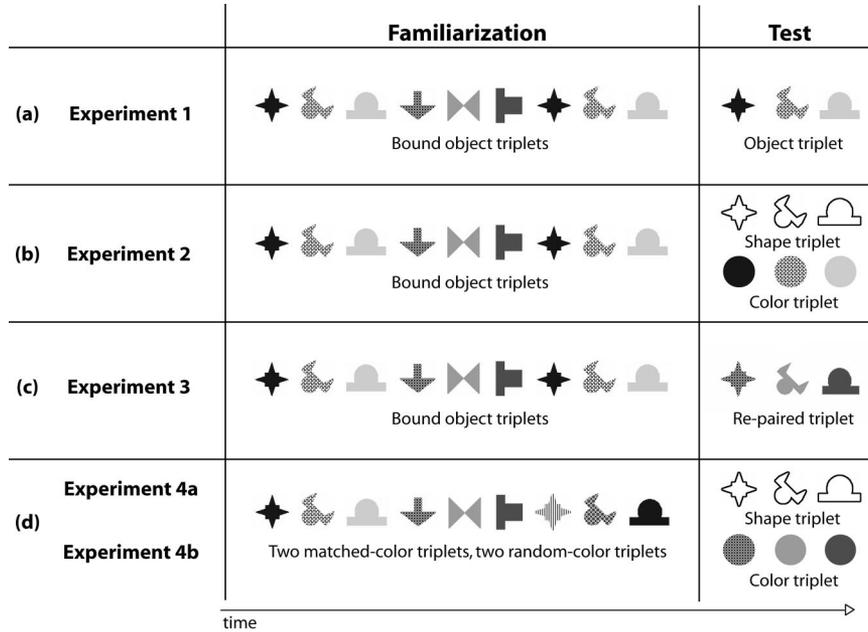


Figure 2. The familiarization and test structure for both shape and color across experiments. See text for details. Color is depicted here in terms of different patterns and gray levels.

performance in Experiment 1 would be high (given that both dimensions were present at test), while average performance across individuals would be low in the feature tests in Experiment 2 (since half of the time observers would have attended to a feature dimension during learning that was not present at test). To explore this possibility, we correlated each observer’s shape performance in Experiment 2 with their color performance in Experiment 2, expecting a negative correlation if observers had attended to only one dimension. In contrast, we obtained a positive correlation ($r = .616$, $df = 14$; $p = .011$), which is consistent only with learning of multidimensional objects.

Experiment 3: Binding Versus Quantity of Statistical Information

The attenuated performance in Experiment 2 relative to Experiment 1 suggests that VSL is object-based, but it remains possible that observers performed better in Experiment 1 because the triplets presented during the test phase itself contained twice as much statistical information (i.e., accumulating over both shape and color dimensions potentially independently), and not because of feature binding between particular shapes and colors. To rule out this explanation, we equated the amount of statistical feature

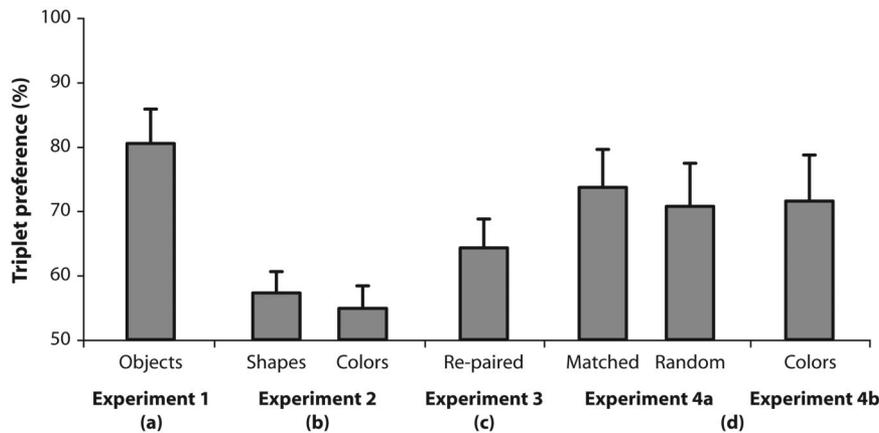


Figure 3. The degree of VSL measured as the percentage of triplets chosen over foil sequences for the bound objects of Experiment 1, the individual feature triplets of Experiment 2, the re-paired feature triplets of Experiment 3, the monochromatic shape triplets of Experiment 4a (as a function of color covariance during familiarization), and the color triplets of Experiment 4b. Error bars depict standard errors. Chance performance is 50%.

information at test in this experiment (such that each test triplet contained both a shape and color triplet) but disrupted the feature binding (such that shape triplets were assigned different color triplets than during familiarization). If performance is still attenuated relative to Experiment 1, we can conclude that VSL operated over bound shape–color pairs and that such differences cannot be explained in terms of the brute amount of statistical information present in the input.

Method

The familiarization phase of this experiment was identical to that of Experiment 1. The test phase, however, pitted “re-paired triplets” against “re-paired foils.” Each re-paired triplet consisted of the shape sequence from one familiarization triplet appearing in the color sequence from a different familiarization triplet. In this way, each shape was assigned to a different color at test than during familiarization (see Figure 2c). These reassignments were kept constant throughout the test and were also used to construct the re-paired foils (which contained unstructured sequences of shape and color). In other words, neither feature of the triplets violated the training sequence, whereas both features of the foils did so.

Results and Discussion

As depicted in Figure 3c, observers reliably preferred the re-paired triplets over the re-paired foils (64%), $t(15) = 3.168$, $p = .006$, $d = 0.792$. Critically, expression of learning was significantly weaker when feature triplets were recombined at test in this experiment versus when presented in the same manner as during familiarization in Experiment 1, $t(30) = 2.327$, $p = .027$, $d = 0.823$. Thus, performance in Experiment 1 cannot solely be attributed to the fact that test triplets contained both shape and color triplets, further supporting the notion that VSL operates over bound shape–color pairs.

Performance was not significantly better in Experiment 3 than in Experiment 2: shape test, $t(30) = 1.253$, $p = .220$, $d = 0.443$; color test, $t(30) = 1.652$, $p = .109$, $d = 0.584$. While our conclusions from this experiment are being driven by statistics rather than by numerical differences, note that our primary conclusion would not be altered even if these differences had reached significance. In other words, the purpose of this experiment was to rule out alternative explanations of Experiment 1 based on the presence of more feature information during test, or on increased surface similarity between familiarization and test phases. Our claim in this respect relies on a comparison only between Experiments 1 and 3. Indeed, the comparison of Experiments 2 and 3 would speak to a different question about whether the amount of feature information at test matters at all. This question seems independent; for example, the number of feature dimensions present at test may modulate performance, while presenting features with the same binding as during familiarization may provide an additional benefit.

Experiment 4a: Feature-Based Learning of Shapes

The first three experiments collectively demonstrate that VSL operates over bound objects, but how general is this conclusion? In

the real world, color and shape do covary for some objects (e.g., bananas) but not others (e.g., t-shirts). Is VSL sensitive to this type of distinction? The previous experiments employed multidimensional objects during familiarization, implemented in a particular way: Both color and shape varied during familiarization and did so in a correlated fashion. However, would VSL still be object-based even if the two feature dimensions did not perfectly covary?

Here, we explored this possibility by testing with monochromatic shapes (as in Experiment 2), but we manipulated the covariance of shape and color during familiarization (Figure 2d): Shapes in two of the triplets (matched-color triplets) were assigned unique colors as in the previous experiments, but the shapes in the remaining two triplets (random-color triplets) were assigned different randomly chosen colors on each presentation.

If VSL is always object-based, then the results of this experiment should parallel those of Experiment 2: The shift to monochromatic shapes should frustrate the expression of learning. However, if VSL is sensitive to the covariance between feature values for individual objects (at the same time that the triplets themselves are being learned), then we might observe object-based learning for the matched-color triplets (i.e., attenuated performance due to the removal of color at test) but feature-based learning for the random-color triplets (i.e., strong performance despite the removal of color at test). A third possibility is that this type of diagnosticity is assessed for entire feature dimensions rather than individual feature values: The presence of at least some shape–color covariance (i.e., in the matched-color triplets) could yield object-based learning for all triplets, or the presence of deviation from such covariance (i.e., in the random-color triplets) could yield feature-based learning for all triplets.

Method

The familiarization phase of this experiment was identical to that of Experiment 2 except as noted here. The two matched-color triplets were sub-sequences of three shapes that were each paired with a unique color throughout the familiarization phase. The two random-color triplets were sub-sequences of three shapes that were paired with three colors drawn randomly on each presentation (from the remaining six colors), with the constraint that no color could appear twice in a row. In other words, shapes in the random triplets could appear in one of six colors at different points during familiarization. In this way, colors in random triplets were nondiagnostic of shape identity, while colors in the matched condition were perfectly diagnostic of shape identity. At test, all shapes were presented monochromatically.

Results and Discussion

As depicted in Figure 3d, observers expressed robust VSL despite the removal of color at test for both the matched-color triplets (74%), $t(15) = 3.973$, $p = .001$, $d = 0.993$; and the random-color triplets (71%), $t(15) = 3.091$, $p = .007$, $d = 0.773$, which did not differ ($t < 1$).

Three key implications follow from these results. First, the lack of any difference between the matched-color and random-color triplets indicates that the diagnosticities of individual color values do not affect the likelihood that shape triplets can be recognized without color. Second, performance for both the matched-color

and random-color triplets was significantly greater than chance—and not reliably different than in the colored-shape test of Experiment 1: matched-color, $t < 1$; random-color, $t(30) = 1.143$, $p = .262$, $d = 0.404$. This suggests that when the covariance between shape and color feature dimensions is disrupted even partially, VSL operates over shapes alone, in a feature-based manner. Third, performance for the (always-color-matched) triplets in Experiment 2 was worse than for the matched-color triplets, $t(30) = 2.403$, $p = .023$, $d = 0.849$; and marginally worse than for the random-color triplets, $t(30) = 1.800$, $p = .082$, $d = 0.636$. This provides further evidence that observers were sensitive to the diagnosticity between shape and color dimensions as a whole, such that color information was discarded—even for the matched triplets—when some colors were nondiagnostic.

Experiment 4b: Feature-Based Learning of Colors

Would this switch to feature-based rather than object-based processing in the face of reduced covariance between feature dimensions also extend to other dimensions, or might it be peculiar to shape processing? Here, we addressed this question by repeating the familiarization phase of Experiment 4a, but now testing the two color triplets without their matched shapes (Figure 2d).

Method

The familiarization phase of this experiment was identical to that of Experiment 4a, with six of the colors (e.g., A, B, C, D, E, F) uniquely paired with one of the shapes from two of the shape triplets, and six of the colors (e.g., G, H, I, J, K, L) randomly presented on each presentation of shapes from the other two shape triplets. The test phase was similar to the color test of Experiment 2, except that there were only two color triplets from familiarization (e.g., ABC, DEF). However, a test of only these two triplets would be underpowered since the two triplets could only be tested against two foils (resulting in eight trials: 2 triplets \times 2 foils \times 2 repetitions). We could not simply use four foils as before because the triplets would then appear twice as often as the foils during test (confounding the familiarity measure): The likelihood of a particular triplet appearing in a test trial would be 50%, while the likelihood of a particular foil appearing would be 25%. To avoid this issue while increasing the number of test trials, we also defined two pseudo-triplets from the six random colors (e.g., HLG, IKJ). The items within the pseudo-triplets were always presented in the same order, similar to the real triplets. The triplets and pseudo-triplets were then used to construct four foil sequences as in the other experiments (e.g., AEL, DLJ, HKA, IBF). Each of the “triplets” was tested against each of the foils twice, resulting in 16 trials containing a real triplet and 16 trials containing a pseudo-triplet. While data from the pseudo-triplet trials were meaningless (since observers had no basis for selecting which pattern was more familiar, as confirmed by chance performance on these trials, $t < 1$), their presence allowed us to equate the frequency of the real triplets with the frequency of all foils during the test phase. Data analysis was then restricted to the test trials containing real triplets.

Results and Discussion

As depicted in Figure 3d, observers expressed robust VSL for the two color triplets despite the removal of shape at test (72%), $t(7) = 2.892$, $p = .023$, $d = 1.023$, thus providing further evidence that VSL can operate in a feature-based manner if the covariance between feature dimensions is disrupted. This conclusion is further supported by the fact that performance was significantly greater than in the color test of Experiment 2, $t(22) = 2.362$, $p = .027$, $d = 0.943$. Moreover, this experiment also provides what we believe is the first demonstration of VSL for colors (or indeed for any feature dimension beyond shape). This rules out the possibility that observers simply discarded color information completely in Experiment 4a in order to learn shape sequences. Instead, low covariance between feature dimensions allowed VSL to operate in parallel in each of these dimensions.

General Discussion

The five experiments reported here constitute the first exploration, to our knowledge, of whether VSL operates over bound objects or lower level visual features. Our results suggest one straightforward implication (that VSL can be object-based), and one surprising possibility (that VSL can perhaps help to determine what counts as an object in the first place). We elaborate on each of these conclusions below.

The Units of VSL

It is perhaps worth noting that we began these studies with no strong prediction about the results. On one hand, many researchers have stressed that VSL can be an automatic low-level process, which operates early in perceptual analysis to provide the initial segmentation on which other later processes depend. As such, we might predict that VSL would operate on those representations which are most common very early in perceptual processing, namely independent visual features. On the other hand, a wealth of data has indicated that discrete bound objects are the currency of many other types of visual processing, including attention (Scholl, 2001), tracking (Alvarez & Scholl, 2005), masking (Moore & Lleras, 2005), and short-term memory (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Xu & Chun, 2006)—among many others. As such, we might predict that VSL would also be fundamentally object-based, operating over representations in which features are already bound together. This possibility would also be consistent with another recent study in which VSL was more robust for sequences that were bound into the same persisting object representation as a result of spatiotemporal continuity (Fiser et al., 2007).

Here we operationalized the question of whether VSL is feature-based by asking whether the expression of VSL would survive the elision of some features at test, while the remaining features still possessed the same statistical structure that they had during familiarization. Our initial results demonstrated that VSL can be object-based: When colors and shapes perfectly covaried (i.e., when each shape had its own unique color) during familiarization, the expression of learning required the presence of both shapes and colors at

test (as in Experiment 1). Of course, there were still robust statistical correlations between the shapes themselves during familiarization, but the expression of learning of these shape sequences was considerably weaker when they were presented in isolation at test (with no reliable expression of learning for colors presented in isolation at test) in Experiment 2. Experiment 3 further demonstrated that the difference in performance between these two experiments was a function of feature binding per se, rather than the mere presence of more feature correlations (due to the added dimensions) present during the test phase of Experiment 1.

The fact that the observers in Experiment 2 largely failed to recognize feature triplets at test is reminiscent of a result from the category learning literature (e.g., Roberts & MacLeod, 1995) showing that categorization cannot be expressed for constituent elements of a rule—analogue to our individual features—when learning is nonstrategic (like VSL, which has been considered a form of incidental learning; see Turk-Browne et al., 2005). Our results also parallel a feature of associative models of Pavlovian conditioning and causal learning (see Williams & Braker, 1999), in which experience with compound cues—analogue to our multidimensional objects—results in associations involving those compounds, but not their constituent elements (e.g., Pearce, 1987; see also Shanks, Charles, Darby, & Azmi, 1998; Williams, Sagness, & McPhee, 1994).

Could VSL Define Objects in the First Place?

The two final experiments (Experiments 4a and 4b) complicated our initial “object-based” conclusion in a particularly interesting way, suggesting that VSL may be object-based only when feature binding is useful—that is, when the values across different dimensions reliably covary. When only half of the shape triplets during familiarization had randomly assigned colors, this effectively disrupted the feature binding seen in the previous experiments (despite the fact that other triplets had perfectly covarying shapes and colors), yielding feature-based VSL—that is, a robust expression of learning when each feature dimension was presented in isolation at test. We could, however, instead characterize VSL as always object-based but interpret our results as indicating that multiple feature values are only combined into single object representations when their covariance is high.

Intuitively, we recognize that some properties are intrinsic to an object, while others are not. We typically think of color as intrinsic: For example, bananas are usually yellow. We do not typically think of luminance as intrinsic to an object, however: A banana may be brighter in the sun and darker in the shade. Why do we treat these two properties differently? On reflection, it seems that the critical difference between these cases is not some special essence of the properties themselves but rather the fact that they are differentially variable with respect to their occurrence in the presence of an object: For example, bananas typically appear in many different lighting situations, but they are usually yellow. We suggest that covariance between surface features in the visual environment serves as an important cue about which object properties are diagnostic (and nondiagnostic). (This point has been made before in the memory literature; see Chalfonte & Johnson, 1996.)

This view can readily accommodate our results. When color and shape are highly correlated (i.e., when all bananas are yellow), then the colors become inextricably bound into the representations that result from VSL, such that removing the colors later on disrupts the expression of VSL (as in Experiment 2a). This disruption may be analogous to ways in which we are slower to recognize purple “bananas” or achromatic fire engines (Naor-Raz, Tarr, & Kersten, 2003; Tanaka & Presnell, 1999). This object-based model could also be applied to the results of Experiments 4a and 4b, where the lack of correlations between features during familiarization may have resulted in separate “object” representations for shapes and colors. In this framework, VSL is always object-based and operates over the largest available set of highly correlated features. This interpretation can be contrasted with the notion discussed in the previous section that VSL becomes feature-based under conditions of reduced feature correlations. Irrespective of the interpretation, however, the sensitivity to correlations between features within objects demonstrated in our experiments may constitute another type of VSL.

Interestingly, the monochromatic shape sequences in Experiment 4a were recognized equally well no matter whether they had been familiarized with perfectly correlated colors or with random colors. Moreover, the only difference between the perfectly correlated shape–color pairs in Experiments 4a and 4b (from matched-color triplets) versus Experiment 2 is that they occurred in the context of other shapes paired with random colors. This demonstrates that observers were sensitive to the correlations between feature dimensions (or possibly the average correlation over all shapes), rather than maintaining separate statistical representations of correlations for individual objects. This effect may be atypical in the broader context of real-world objects, though, where color diagnosticity differs from object to object (e.g., consider a fire engine vs. a lamp; Tanaka & Presnell, 1999) and may be moderated by the degree to which feature dimensions are processed in a separable or integral manner (e.g., Nosofsky & Palmeri, 1996; see also Garner, 1974). Also, children seem to be reluctant at some ages to categorize objects by color (despite intact color perception), but they will do so robustly for some items that have reliable colors, such as foods (Macario, 1991).

Sensitivity to feature correlations may in fact hold more generally, explaining why other types of statistical regularities are or are not learned. A study of spatial VSL, for example, demonstrated that small spatial groups of larger spatial arrays were learned only when the smaller groups could vary in their spatial relationships to each other (Fiser & Aslin, 2005). When the small “parts” were always arranged into the same global patterns, in contrast, observers were not later sensitive to the intra-part statistical structures when they were tested by themselves. In other words, testing the local parts later in isolation (from the other local parts with which they had been previously associated) was equivalent to testing the shapes in our experiments in isolation from the colors with which they had been previously associated. Thus, statistical structure may serve to define objects in terms of both spatial scale (as argued by Fiser & Aslin, 2005) and surface features (as suggested here). Fiser and Aslin (2005) argued that their results reflect a constraint on “embeddedness” in hierarchical structures, which essentially serve to bind objects into global scenes. Our results, in a complementary fashion, indicate that statistical

covariance can also serve to determine which surface features are bound into objects themselves.⁵

In fact, the idea that associative variability may serve to define the underlying units of processing has a long history, going back to William James: “What is associated now with one thing and now with another, tends to become dissociated from either, and grow into an object of abstract contemplation by the mind” (James, 1890, p. 506). This concept has also been important in the long-term memory literature. According to the encoding variability hypothesis (Martin, 1968), for example, the more often a stimulus is encountered within a particular context, the more deeply entrenched task-relevant attributes of the stimulus become. Relative to an item that has been encountered an equal number of times in multiple contexts, memory for the fixed-context item will be impaired in a novel context (both because it is associated with fewer retrieval cues and because the associations to these retrieval cues are stronger). In Experiment 2, triplets of black (novel color context) shapes may have been unrecognizable because of the fixed color context in which they were encountered during familiarization. Conversely, in Experiment 4a where some shapes were encountered in multiple color contexts, recognition of black shape triplets was spared. Where this account differs from the sensitivity to covariance demonstrated here is in terms of automaticity: Conventional demonstrations of encoding variability have defined context as the task in which subjects were engaged (e.g., categorization vs. recognition; Wagner, Maril, & Schacter, 2000), while context in the current study was based on surface features themselves.

Conclusions

We began this project in an attempt to determine the impact of features vs. objects on VSL, and in the process we discovered that in some cases VSL can be object-based. In the end, however, our results and their connections to other studies suggest that statistical learning may also serve to construct objects out of some features but not others. In this sense, statistical learning may reflect not a unitary process but rather a hierarchy of interacting processes, some of which help to determine the input to others. Thus, whereas VSL has often been studied in an isolated way, it may actually be an integral component of many other more familiar types of visual processing.

⁵ In both the present experiments and the studies of Fiser and Aslin (2005), smaller features (shapes or parts) may be learned despite the presence of other statistical structure (in colors or more global hierarchies) when there is high variability. However, we have previously shown that statistical regularities in shape sequences are learned when there is especially low color variability (such that many different shapes can appear in the same color; Turk-Browne et al., 2005). Onnis and Christiansen and their colleagues demonstrated a similar effect in a very different context: Nonadjacent dependencies between syllables can be learned when the intervening syllable(s) have extremely high or low variance, but not mid-dling variance (Onnis, Christiansen, Chater, & Gómez, 2003; Onnis, Monaghan, Christiansen, & Chater, 2004).

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111.
- Alvarez, G. A., & Scholl, B. J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General, 134*, 461–476.
- Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15*, 460–466.
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403–416.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28–71.
- Comtois, R. (2004). VisionShell PPC [Software libraries]. Cambridge, MA: Author.
- Conway, C. M., & 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.
- 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. (2002a). 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. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 15822–15826.
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134*, 521–537.
- Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review, 14*, 173–178.
- Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- Harris, Z. (1955). From phoneme to morpheme. *Language: Journal of the Linguistic Society of America, 31*, 190–222.
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*, 658–680.
- James, W. (1890). *The principles of psychology*. London: Macmillan.
- Jungé, J., Turk-Browne, N. B., & Scholl, B. J. (2005). Visual statistical learning through intervening noise [Abstract]. *Journal of Vision, 5*, 421a.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.
- Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development, 6*, 17–46.
- Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review, 75*, 421–441.
- Moore, C. M., & Lleras, A. (2005). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1171–1180.
- Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32*, 667–680.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1–32.
- Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review, 3*, 222–226.
- Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual

- attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1299–1313.
- Onnis, L., Christiansen, M.H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886–891). Mahwah, NJ: Erlbaum.
- Onnis, L., Monaghan, P., Christiansen, M.H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1047–1052). Mahwah, NJ: Erlbaum.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Roberts, P. L., & MacLeod, C. (1995). Representational consequences of two modes of learning. *The Quarterly Journal of Experimental Psychology*, 48A, 296–319.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80, 1–46.
- Shanks, D. R., Charles, D., Darby, R. J., & Azmi, A. (1998). Configural processes in human associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1353–1378.
- Stadler, M. A., & Frensch, P. A. (1998). *Handbook of implicit learning*. Thousand Oaks, CA: Sage.
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, 61, 1140–1153.
- Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.
- Turk-Browne, N. B., Johnson, M. K., Chun, M. M., & Scholl, B. J. (2007). *Neural evidence of statistical learning: Efficient detection and anticipation of regularities without awareness*. Manuscript submitted for publication.
- Turk-Browne, N. B., Jungé, J., & 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. (2006). The space–time continuum: Spatial visual statistical learning produces temporal processing advantages [Abstract]. *Journal of Vision*, 6, 676a.
- Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, 12, 52–60.
- Williams, D. A., & Braker, D. S. (1999). Influence of past experience on the coding of compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 461–474.
- Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 694–709.
- Xu, Y., & Chun, M. M. (2006, March 2). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440, 91–95.

Received April 18, 2006

Revision received November 5, 2007

Accepted November 7, 2007 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.