VISUAL COGNITION, 2011, 19 (4), 469–482

Psychology Press
Taylor & Francis Group

Visual learning of statistical relations among nonadjacent features: Evidence for structural encoding

Elan Barenholtz[1] and Michael J. Tarr[2]

[1]Department of Psychology, Florida Atlantic University, Boca Raton, FL, USA
[2]Center for the Neural Basis of Cognition, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Recent results suggest that observers can learn, unsupervised, the co-occurrence of independent shape features in viewed patterns (e.g., Fiser & Aslin, 2001). A critical question with regard to these findings is whether learning is driven by a structural, rule-based encoding of spatial relations between distinct features or by a pictorial, template-like encoding, in which spatial configurations of features are embedded in a "holistic" fashion. In two experiments, we test whether observers can learn combinations of features when the paired features are separated by an intervening spatial "gap", in which other, unrelated features can appear. This manipulation both increases task difficulty and makes it less likely that the feature combinations are encoded simply as larger unitary features. Observers exhibited learning consistent with earlier studies, suggesting that unsupervised learning of compositional structure is based on the explicit encoding of spatial relations between separable visual features. More generally, these results provide support for compositional structure in visual representation.

Keywords: Perceptual learning; Statistical learning; Vision.

How are the spatial relations between individual visual features learned from complex visual input, such as natural scenes, for later recognition? A number of studies have demonstrated that observers can learn statistical contingencies among disparate visual features in unsupervised tasks (Fiser & Aslin, 2001, 2005; Orban, Fiser, Aslin, & Lengyel, 2008). In such experiments,

Please address all correspondence to Elan Barenholtz, 777 Glades Road, Boca Raton, FL, USA. E-mail: elan.Barenholtz@fau.edu

This research was supported in part by NGA Award No. HM1582-04-C-0051 to EB and MJT, an NIH EUREKA Award No. 1R01MH084195-01 to MJT, and by an NSF Award No. BCS-0958615 to EB.

http://www.psypress.com/viscog DOI: 10.1080/13506285.2011.552894

observers viewed patterns made of multiple shape features arranged in a grid. Unknown to the observers at the initiation of the experiment, the features making up the grid patterns consisted of "base" pairs or triples of features that always appeared together in the same spatial configuration. Observers in these experiments showed an ability to identify the base combinations as belonging together in subsequent testing.

A central question that applies to such studies is whether learning the relations between features is driven by a holistic, "pictorial" form of learning or a rule-based, "structural" form of learning, a dichotomy at the root of different theoretical approaches to object recognition. In a holistic representation, encoding and recognition proceed on the basis of comparing some presented pattern to an image "whole-cloth" without extracting features or their relations—essentially a template. This approach is represented in template-based models of visual recognition such as "Recognition-By-Alignment" (Ullman, 1989), in which a stored representation is matched to a new example on the basis of a global transformation of the entire image. As one alternative, consider a "feature-based" approach in which locally defined properties are extracted from the image without any regard to their spatial configuration with respect to one another, that is, as a "bag of features" (Mel, 1997). Finally, a "structural" approach is based on a representation in which independent features *and* relations are explicitly encoded. According to this family of models, recognition proceeds on the basis of first identifying a set of features, as well as determining whether these features are in the appropriate relations relative to one another. An example of this form of recognition is reading, in which the evidence suggests that individual letters must first be identified in order for the word that they comprise to be recognized (Pelli, Farell, & Moore, 2003).

This latter, structural, approach is represented by a number of influential models of object recognition (e.g., Biederman, 1987; Marr & Nishihara, 1978) in which features and their relations are both represented explicitly. Although such models of recognition have been extremely influential from a theoretical perspective, empirical research conclusively supporting the structural aspects of such approaches has been more difficult to establish. This is due in part to the fact that any configuration of features may, in theory, be encoded on the basis of a larger, "holistic" feature that encompasses the smaller ones (for an extensive discussion of this issue, see Barenholtz & Tarr, 2007). A good example of this ambiguity is provided by studies of face recognition that aim to compare the role of individual features of a face—such as the eyes and mouth—and their configuration—such as the spacing between the eyes (Maurer, le Grand, & Mondloch, 2002; Tanaka & Farah, 1993). In such studies, any given configuration of features may be captured by using larger, "holistic" facial features—for example a template consisting of the entire image in which the eyes, as well as the space between

them—rather than explicitly encoding configural information. Indeed several models of face recognition explicitly use such "larger features" in lieu of explicitly encoding relations (Nestor, Vettel, & Tarr, 2008; Zhang & Cottrell, 2005).

All of the previously discussed studies of statistical visual pattern learning (e.g., Fiser & Aslin, 2001) contain a similar ambiguity with regard to the underlying learning mechanism. That is, subjects in these studies could be learning a structural "rule" that Feature X and Feature Y always co-occur in a specific spatial configuration. To do so would require an "explicit" encoding of the relation between the features.[1] Alternatively, subjects could be learning an image template that encompasses both Feature X and Feature Y—as well as the intervening space between them. In this case, the relation between the features would be "implicit"; indeed the individual features themselves may not even be separately encoded in this type of representation. In particular, the latter is a plausible explanation for these earlier studies because the paired features were always spatially adjacent to one another. Thus, even though the patterns appear to be made up of different features separated by white space and grid lines, it is possible, and perhaps likely, that the repetition of the exact same image (e.g., the image containing Feature X and Feature Y in a specific configuration) leads to an encoding of the entire pattern as a whole, a.k.a. a template. Indeed, one result described by Fiser and Aslin (2005) seems to potentially favour a holistic account in which the individual features of the learned patterns are not instantiated as separable entities (i.e., they are not "compositional"). In Fiser and Aslin's study, pairs of feature that appeared as part of larger combinations (such as triplets) were not learned as well as pairs that did not appear as a subset of a larger configuration. This seems to be more readily explained by a holistic interpretation, in which the repeated triplet patterns were encoded as a unitary image and that a subfeature of this pattern did not register a match. A structural account would have more difficulty explaining these results since it is not possible to encode the relation among three individual features without also encoding the relations between pairs of features making up the triplet.

In order to more specifically address whether structural encoding takes place when learning feature relations from complex patterns, our study introduced a learning task similar to Fiser and Aslin's (2001), but one in which members of base–feature pairs were spatially nonadjacent—they were separated by an intervening "gap" in which other, unrelated features appeared. The introduction of an intervening gap itself makes the task more

[1] We use the terms "explicit" and "implicit" to refer to whether a property is independently defined within the learned representation; we are not referring to another meaning of these terms within the memory literature where their senses refer to whether people are consciously aware of their learning or not.

difficult, simply because the to-be-learned relations occur across a larger visual distance. More importantly, we reasoned that the presence of an intervening space encompassing variable features between consistent features makes it unlikely that the base pairs will be encoded as single "holistic" features, that is, encoding a single image that is repeated consistently (as in earlier studies). Instead, learning feature pairs is presumed to require a structural encoding based on the spatial relationship between distinct features.

EXPERIMENT 1

Methods

Subjects. Fifteen Brown University undergraduate students participated in exchange for cash payment.

Stimuli and procedure. Each subject performed two separate blocks of trials: a learning phase and a test phase.

Stimuli in the learning phase consisted of polygonal shapes ("features") arranged in a 4×4 grid of possible locations (Figure 1). Each of the polygonal shapes had a unique colour.[2] Unlike the studies discussed earlier, in our study the grid was not visible. Each of the individual shapes measured approximately 1.3×1.3 degrees of visual angle and consisted of a bilaterally symmetrical polygon with a distinct colour. Of the 22 shapes available, a subset of 16 was chosen and assigned to one of three categories: vertical pairs (six shapes comprising three pairs), horizontal pairs (six shapes comprising three pairs), and singletons (four shapes). The members of each vertical pair always appeared together within a pattern and were always in a configuration in which one of the features from the pair (the same one on every trial) was two grid spaces above the other, horizontally aligned, within the grid. Similarly, the horizontal pairs always appeared together in a configuration in which one of the features was two grid spaces to the right of the other, vertically aligned, within the grid. On any given trial, either two or three pairs appeared in the overall pattern, consisting of either one horizontal pair and one vertical pair, two horizontal pairs and one vertical pair, or two vertical pairs and one horizontal pair. The specific choice of

[2] Previous research in statistical learning has found that when colour and shape properties are constantly coupled (i.e., when a specific shape always appears paired with a specific colour), statistical learning operates over the two dimensions as a single "object", not based on either property independently (Turke-Brown, Isola, Scholl, & Treat, 2008). Thus, we refer to these colour/shape-based units as individual "features". Colour was included in our experiments in order to make the task easier by making the individual features more dissimilar from one another; in particular, in an attempt to offset the greater difficulty inherent in learning relations among spatially distant features.

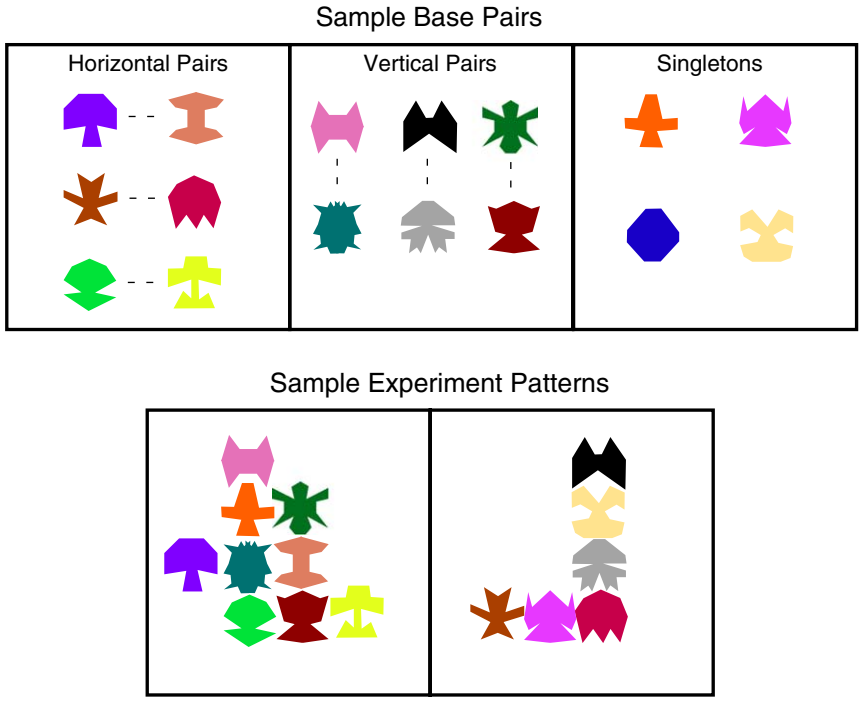

Figure 1. Example stimuli used in Experiments 1 and 2. Top: The stimuli consisted of patterns of vertical and horizontal "base pairs", which always appeared together in the same configuration, as well as singletons which appeared in the gap between base-pair members whenever that location was not occupied by a feature from another pair. Bottom: Two example stimulus patterns made up of the base pairs and singletons shown above. [To view this figure in colour, please visit the online version of this Journal.]

pairs was randomized across trials. Note that no constraint was made on where the pairs appeared within the virtual 4×4 grid except that no features could occupy the same location. As can be seen in the bottom of Figure 1, this means that different patterns could span more or less of the grid than others. Singletons did not have any specific relation to any of the individual paired shapes but appeared within the space between the vertical or horizontal pairs on any trial when that space was otherwise empty (i.e., when no member of the other pairs appeared within that space). This ensured that the base pairs were always separated by another, unrelated, feature. The total number of features on screen for any pattern thus ranged from a minimum of four (two pairs with no singletons) to a maximum of nine (three pairs with three singletons).

The learning phase consisted of an unsupervised learning task—that is, subjects were never given any information regarding the statistical structure

of the patterns that they were intended to learn. However, in order to encourage subjects to attend to the details of the patterns, we employed a simple change detection in which subjects had to determine whether two sequential patterns were the same or different on each trial. (This may be contrasted with earlier studies of statistical learning in which patterns were presented sequentially without interruption). The experimental sequence was as follows: On each trial a single pattern appeared in a random location on the screen within 4 degrees horizontally and vertically offset from the centre of the screen. After 2 s, the pattern disappeared, followed by a grey screen for 1 s. Then a second pattern appeared in a new random location for 1 s, after which it disappeared. This change in the location of the grid was intended to eliminate a relation between the absolute screen position and the shape features; that is, encouraging subjects to learn the relative positions of the features in the grid. On half of the trials ("no-change trials"), the second pattern was identical to the first one; in the other half ("change trails"), either one of the base pairs or one of the singletons was switched for another pair or singleton that had not been shown in the first pattern (see Figure 2 for examples of the two kinds of changes). To ensure that subjects needed to attend to the individual features, and not just the global shape of the pattern, the new pair or singleton introduced on change trials was always in the same location as the one it replaced. Note that the change never resulted in a member of a base pair appearing with some other feature in the paired position other than its respective pair. After the second pattern disappeared, the subject was prompted to respond whether the two patterns were the same or different. This task was unsupervised—no feedback was given for correct or incorrect responses. Each subject performed 300 learning trials.

In the test phase, subjects discriminated between true base pairs from the learning phase and "nonpairs": pairs of features selected from two different base pairs that were conjoined in either a horizontal or vertical orientation for the test phase only. Note that the exact same features were used to construct the nonpairs as the original learned base-pairs; the only difference was in the pairing of the features. When applicable, the left/right and top/bottom features of the nonpairs were shown in the same relative position of the configuration in which they had appeared during the learning phase. For example, a feature that was the left member of a horizontal pair during learning always appeared on the left side when it was part of a horizontal nonpair stimulus. This constraint ensured that subjects could not perform the test task based on the fact that certain features tended to be on the right/left or top/bottom of the overall pattern, since both pairs *and* nonpairs preserved this property. Instead, subjects needed to learn which other feature each feature was paired with. There were a total of six nonpairs—three horizontally configured and three vertically configured—to match the six

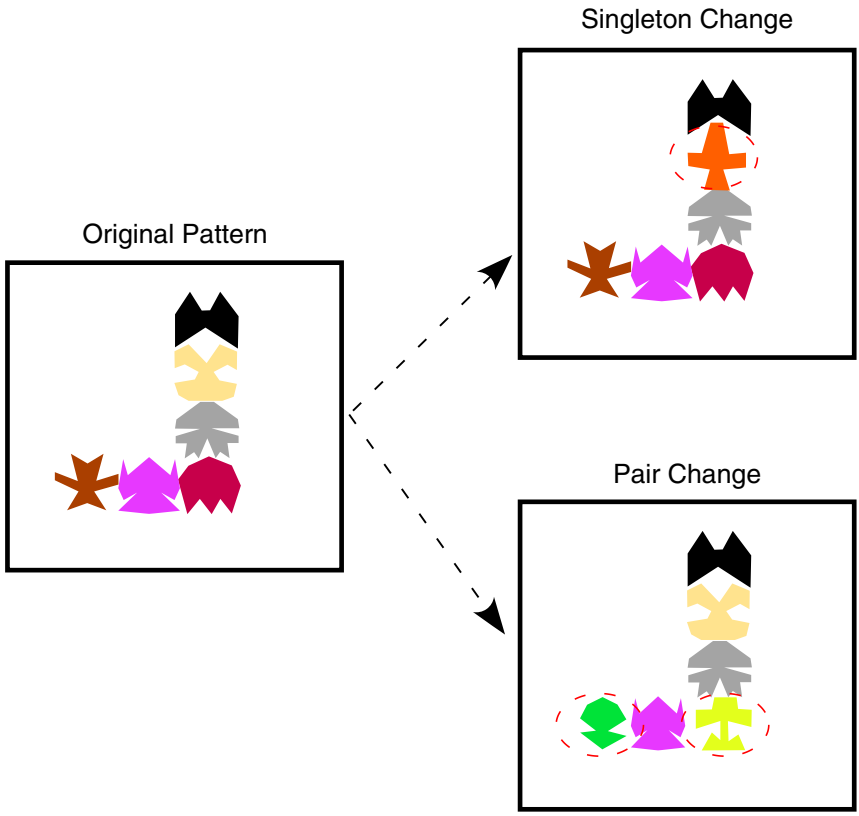

Figure 2. Example of the two types of changes used in the learning phase change detection task (see text). On the left is the original pattern and on the right are the changed versions, with the altered features circled in red. [To view this figure in colour, please visit the online version of this Journal.]

true base-pairs. The test sequence was as follows (Figure 3): On each trial, subjects saw a single pair of features—consisting of two features shown in either a horizontal or vertical configuration, without a singleton between them—for 2 s. This was followed by a grey screen for 1 s and then another pair of features for 2 s. On half of the trials, the true base pair was shown first, followed by the nonpair; on the other half of the trials this order was reversed. After the two pairs had been shown, subject were asked to determine which of the two displayed pairs, the first or the second, looked more familiar (based on the patterns they had viewed in the learning phase). Each subject saw the six base pairs compared to each of the six nonpairs for a total of 36 trials per subject.

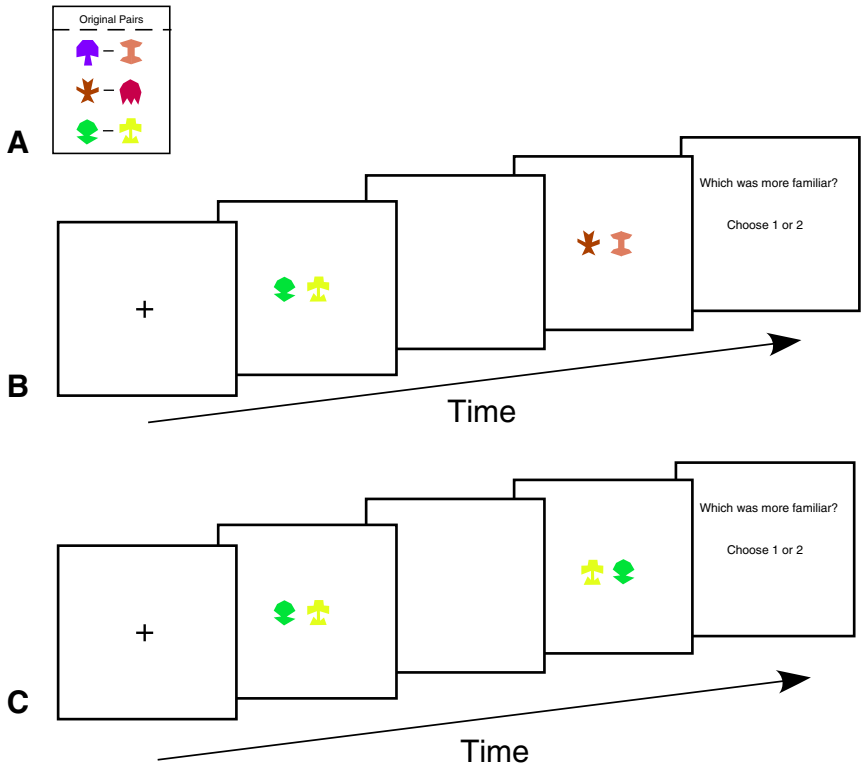

Figure 3. Schematic diagrams of a test phase trial from Experiments 1 and 2. (A) Examples of original base pairs. (B) A test sequence from Experiment 1 in which a pair and a nonpair are shown in succession. Subjects had to choose which was more familiar based on the learning phase. (C) A test sequence from Experiment 2 in which a pair and its reversal are shown in succession. Subjects had to choose which was more familiar based on the learning phase. [To view this figure in colour, please visit the online version of this Journal.]

Results and discussion

The overall accuracy in the learning phase was 78%. In the test phase, we calculated the percentage of trials on which subjects chose, on average, the base pair as being more familiar than the nonpair. There was a tendency to choose the base pairs as being more familiar than the nonpairs (mean = 62%, $SE = 2\%$; range 47–72%). A t-test comparing observed performance to hypothetical chance performance found that this tendency was significant, $t = 3.267$, $p < .001$. This level of performance is comparable to earlier studies of statistical learning (e.g., Exp. 2 of Fiser & Aslin, 2001). On the one hand, this is somewhat surprising since the task in our study would appear to be much more difficult as it requires that subjects encode relations among spatially distant features. One possible advantage in our

experiment, however, compared with these earlier studies, is that the features contained colour information, whereas earlier studies used black and white stimuli, making any direct comparison between performance levels difficult.

In postexperiment interviews, we asked subjects whether they had noticed that certain features appeared in pairs during the learning phase. Only one subject reported explicitly noticing the paired stimuli; most subjects reported being unaware of any statistical structure during the learning phase. These results are consistent with recent reports concerning the "implicit" nature of statistical learning (Kim, Seitz, Feenstra, & Shams, 2009).

EXPERIMENT 2

Because the members of a base pair always appeared together and always in the same configuration, there are two possible sources of information that subjects might have used to distinguish between base pairs and nonpairs in the test task: the constant conjunction of the paired features (i.e., A always appears with B) and their configuration (i.e., A always appears to the right of B). To determine whether subjects encoded configural information, and not just the constant conjunction, Experiment 2 had subjects perform the same learning task but with a test task in which they had to choose between learned and novel configurations of the same features.

Methods

Ten Brown University undergraduates who did not participate in Experiment 1 and were naïve to the purpose of the study participated for class credit.

Stimuli and procedure. The stimuli and procedure were identical to Experiment 1 except for the nature of the comparison task in the test phase. On each trial of the test phase in Experiment 2, subjects had to choose between base pairs and "reverse pairs", consisting of the same features as the base pair but in reverse configuration (i.e., if the base pair consisted of A shown to the right of B during learning, the reverse pair consisted of A shown to the left of B) (Figure 3B). As in Experiment 1, each of the true base pairs was compared with each of the reverse base pairs for a total of 36 trials per subject.

Results and discussion

The overall accuracy in the learning phase was 82%. In the test phase there was a tendency for subjects to chose the base pairs as more familiar than the reverse pairs (mean = 66%, $SE = 3\%$, range 41–78%). A t-test comparing

observed performance to hypothetical chance performance found that this tendency was significant, $t(10) = 5.29$, $p < .0001$. These results demonstrate that subjects learned the spatial configuration of the members of the pairs, not just their constant conjunction

GENERAL DISCUSSION

Our study extends earlier experiments showing unsupervised learning of relations between distinct visual features to cases in which the related features are spatially nonadjacent. Unlike earlier studies, in our experiments the learned pairs did not constitute a single, consistently repeated image, which could serve as a template to be matched directly to a test stimulus. This makes a holistic "pictorial" representation much less likely, although not impossible. For example, one possible way of doing the task using a simple template-based system would be to determine that the correct test pair is *more* globally similar—though not identical—to the images presented during learning. However, without highlighting the stable pairs, there is no means for establishing which images are "important" during learning, effectively requiring that *all* of the presented patterns be considered equally. A more plausible interpretation is that learning depends on somehow detecting the stable relations among features and encoding that information while ignoring the arbitrarily intervening features. Such an encoding scheme, which depends on detecting relations among spatially distant features, is inherently structural.

Of course, the evidence for structural learning in the current study does not rule out the possibility that nonstructural encoding may take place under other circumstances, including previous experiments of statistical learning. In particular, as mentioned earlier, the finding by Fiser and Aslin (2005) that pairs of features that were members of triplets were not learned separately from the larger configuration seems to favour a holistic account. One possibility is that patterns made up of adjacent features—as in all previous studies—may be learned holistically while patterns consisting of nonadjacent pairs (as in the current study) are learned structurally. Thus, it is possible that statistical learning takes on different forms depending on the nature of the statistical information itself; where applicable, a template-like system might be a more efficient learning mechanism. However, a structural mechanism might be employed by observers to extract statistical information where variability in the stimuli—such as in our study—requires it.

Thus, the current study provides evidence for structural encoding, as well as offering a good example of why it might be useful. In short, by using discrete features as well as spatial relations, structural models provide the ability to generalize across different examples, while retaining specificity. For

example, identifying a "smiley face" on the basis of a collection of features (e.g., two circles and a curved arc), rather than an entire image of a smiley face, allows identification across examples in which the spacing between the features varies. However, the potential downside of a pure feature-based model is that the chosen features may be too generic and therefore underconstrain the model, leading to false positive matches. Encoding relations along with the features as in structural models—e.g., the two circles must be aligned and above the arc to be a smiley face—can serve to overcome this problem. In the current study, subjects were able to identify the true base pairs, despite the variability across different examples of the pairs (Experiment 1) while still encoding enough specificity to correctly identify the base pairs from among distractors that shared the same features (Experiment 2).

As noted, the potential utility of structural encoding has long been recognized and several influential theories of visual object recognition have been explicitly structural (Biederman, 1987). On the empirical side, a number of experimental studies have looked directly at the role of relations in the encoding of complex objects (Hummel & Stankiewicz, 1996; Saiki & Hummel, 1998), specifically as it concerns object recognition. However, many of these studies contain the same ambiguity concerning "structural" or "holistic" encoding discussed earlier. For example, Arguin and Saumier (2004) compared visual search times for when the search target and distractors shared either features (volumetric parts) or configurations (the spatial organization of the parts). They found that these two properties—parts and relations—contributed independently to search times. Similarly, Keane, Hayward, and Burke (2003) used a change detection task to assess the perceptual similarity of objects that had shared parts in different configurations versus shared configurations of different parts, finding that shared configurations yielded higher perceptual similarity (i.e., worse change detection). However, in both of these studies, objects that shared configurations of different parts also shared similar global shapes. Thus, it is possible that subjects in these experiments did not explicitly encode the configural information between separate parts per se. The current study, while providing perhaps some of the strongest evidence to date for structural encoding, does not explicitly address the role of structure in typical applied-recognition tasks such as face and object recognition. Thus, further studies will be needed to address the scope and application of this type of learning.

Whereas the current study is concerned with statistical learning in static visual patterns—a condition we believe is particularly relevant to the topic of visual recognition—several previous studies have examined statistical learning of *temporal* patterns, in serially presented sequences of visual or auditory stimuli. For example, a number of studies have found that subjects can learn to identify "pairs" of visual features, similar to those used here, where the

pair is defined by one feature always being presented immediately before or after the other feature in a serially presented sequence (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Jungé, & Scholl, 2005). Using this methodology, a recent study by Turk-Browne and Scholl (2009) found an interesting correlate to Experiment 2 of the current study; in addition to identifying which features had been paired, subjects in their experiment were able to discriminate the correct *order* in which they had been presented. It is worth noting that the pairs in these studies are always "adjacent" in time—i.e., no additional features appear intermediately between paired features—and thus may also be encoded "holistically" in terms of a rigid temporal pattern, much like the stimuli in earlier studies of static patterns discussed previously. However, a number of studies of temporal statistical learning using *auditory* stimuli have demonstrated that people –as well as nonhuman primates—can learn the relations between pairs of *nonadjacent* features, i.e., even when the paired sounds were separated by an intervening, unrelated sound (Creel, Newport, & Aslin, 2004; Gebhart, Newport, & Aslin, 2009; Newport, Hauser, Spaepen, & Aslin, 2004). Such findings suggest interesting parallels to the results of our present study. However, it is impossible to say at this point whether these superficially similar results reflect a more general learning mechanism concerned with relations between distinct features—regardless of the perceptual dimension—or whether they reflect distinct mechanisms particular to each different dimension (in this case time and space). In particular, learning sequences of both visual and auditory signals depends on cognitive systems devoted to temporal encoding, even perhaps representing specific linguistic or prelinguistic mechanisms. In contrast, visual pattern learning of the sort described here depends on cognitive systems devoted to spatial encoding and, we would argue, visual recognition. Thus, there may be little overlap in the underlying mechanisms supporting the learning of nonadjacent features in different perceptual domains. At the same time, the possibility that such learning depends on a more general, shared statistical mechanisms is an intriguing one that merits further research.

REFERENCES

Arguin, M., & Saumier, D. (2004). Independent processing of parts and of their spatial organization in complex visual objects. *Psychological Science*, *15*(9), 629–633.

Barenholtz, E., & Tarr, M. J. (2007). Reconsidering the role of structure in vision. In A. B. Markman & B. H. Ross (Eds.), *The psychology of learning and motivation: Advances in research and theory. Vol. 47: Categories in use*. New York, NY: Academic Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130.

Fiser, J. Z., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Fiser, J. Z., & 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.

Fiser, J. Z., & Aslin, R. N. (2005). Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review*, *16*(3), 486–490.

Hummel, J. E., & Stankiewicz, B. J. (1996). Categorical relations in shape perception. *Spatial Vision*, *10*, 201–236.

Keane, S. K., Hayward, W. G., & Burke, D. (2003). Detection of three types of changes to novel objects. *Visual Cognition*, *10*(1), 101–127.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*, 145–149.

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

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of spatial-organization of 3-dimensional shapes. *Proceedings of the Royal Society of London: Biological Sciences*, *200B*(1140), 269–294.

Maurer, D., le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

Mel, B. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, *9*, 977–804.

Nestor, A., Vettel, J. M., & Tarr, M. J. (2008). Task-specific codes for face recognition: How they shape the neural representation of features for detection and individuation. *PLoS One*, *3*(12).

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117.

Orban, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the USA*, *105*(7), 2745–2750.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, *423*(6941), 752–756.

Saiki, J., & Hummel, J. E. (1998). Connectedness and part-relation integration in shape category learning. *Memory and Cognition*, *26*, 1138–1156.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 399–407.

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. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 195–202.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*(3), 193–254.

Zhang, L., & Cottrell, G. W. (2005). Holistic processing develops because it is good. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual Cognitive Science conference* (pp. 2428–2433). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Manuscript received May 2010
Manuscript accepted December 2010