Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 1, 195–202

0096-1523/09/$12.00 DOI: 10.1037/0096-1523.35.1.195

Flexible Visual Statistical Learning: Transfer Across Space and Time

Nicholas B. Turk-Browne and Brian J. Scholl
Yale University

The environment contains considerable information that is distributed across space and time, and the visual system is remarkably sensitive to such information via the operation of visual statistical learning (VSL). Previous VSL studies have focused on establishing what kinds of statistical relationships can be learned but have not fully explored how this knowledge is then represented in the mind. These representations could faithfully reflect the details of the learning context, but they could also be generalized in various ways. This was studied by testing how VSL transfers across changes between learning and test, and the results revealed a substantial degree of generalization. Learning of statistically defined temporal sequences was expressed in static spatial configurations, and learning of statistically defined spatial configurations facilitated detection performance in temporal streams. Learning of temporal sequences even transferred to reversed temporal orders during test when accurate performance did not depend on order, per se. These types of transfer imply that VSL can result in flexible representations, which may in turn allow VSL to function in ever-changing natural environments.

Keywords: statistical learning, generalization, transfer, spatial layout, temporal order

A fundamental aspect of visual perception is that the input at any given moment (and in any given spatial location) is ambiguous. In order to resolve this ambiguity, the visual system must therefore take advantage of past experience both during extensive ontogenetic training (perceptual learning; e.g., Fahle & Poggio, 2002) and over phylogenetic time (resulting in hardwired visual heuristics; e.g., Marr, 1982). Such learning may then result in efficient "unconscious inferences" about the sources of ambiguous input (von Helmholtz, 1910/1925). On more local timescales, learning and past experience may also enhance visual performance, for example, by speeding online processing. One specific type of learning—visual statistical learning (VSL)—may be particularly important in this context, facilitating response times in visual search of spatial layouts and temporal sequences (e.g., Baker, Olson, & Behrmann, 2004; Chun & Jiang, 1998; Hunt & Aslin, 2001; Olson & Chun, 2001; Turk-Browne, Jungé, & Scholl, 2005).

Previous studies of VSL have largely focused on the types of statistical relationships that can be learned, including spatial configurations (e.g., Chun & Jiang, 1998; Fiser & Aslin, 2001, 2005), temporal sequences (e.g., Baldwin, Andersson, Saffran, & Meyer, 2008; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Olson & Chun, 2001; Turk-Browne et al., 2005; Turk-Browne, Johnson, Chun, & Scholl, 2008), nonadjacent dependencies (e.g., Turk-Browne et al., 2005), and covariance among visual surface features (e.g., Turk-Browne, Isola, Scholl, & Treat, 2008). Existing research, however, has had relatively little to report about what exactly is learned as a result of VSL. It is possible that the mental representations generated by VSL are highly specific to all details of the learning context, such that learning can only be later expressed in the same context, but this need not be the case. Learning that could only be expressed during later exposure to identical input patterns would be of little use in natural visual environments wherein contexts (and observers) are constantly changing. As a result, VSL may involve generalization from the input, such that the resulting representations can be more flexibly applied later to a variety of inputs.

In fact, results from neighboring literatures suggest that VSL may not involve such generalization. Perhaps the central result of the large literature on perceptual learning is the remarkable degree of specificity in the resulting visual representations. For example, extended training in a visual discrimination task results in representations that are specific to the retinotopic location, orientation, stimulus, task, and context from training, such that learning fails to transfer across changes in these dimensions (e.g., Crist, Kapadia, Westheimer, & Gilbert, 1997). Researchers have obtained similar results using paradigms closer to the types of statistical learning studied here. For example, learning that a particular spatial configuration predicts the location of a target during visual search does not transfer to an identical configuration composed of items of a different color (a type of "hyperspecificity" of contextual cuing; Jiang & Song, 2005). Thus, in the visual domain, specificity is commonplace.

To our knowledge, there have been no published studies of this type of generalization in the study of VSL per se, wherein an

---

Nicholas B. Turk-Browne was supported by a Foreign Natural Sciences and Engineering Research Council of Canada postgraduate scholarship. Brian J. Scholl was supported by National Science Foundation Grant BCS-0132444. Parts of this work were presented at the 6th Annual Meeting of the Vision Sciences Society in Sarasota, Florida (May, 2006) and the 47th Annual Meeting of the Psychonomic Society in Houston, Texas (November, 2006). We thank Luca Bonatti, Tim Brady, Phillip Isola, Christian Luhmann, and the members of the Scholl and Chun labs at Yale University for helpful conversation and comments on drafts of this article. We thank Dick Aslin for providing the shapes used in these studies, which we also used in Turk-Browne, Jungé, and Scholl (2005).

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

undifferentiated temporal stream or spatial layout is parsed into units only via the extraction of statistical regularities.[1] Generalization and abstraction have been observed in related forms of learning, however. In studies of artificial grammar learning (AGL), for example, stimulus sequences are explicitly parsed into discrete units, but underlying grammars define the structure of each individual unit. These learned grammars can then transfer to new elements (e.g., Reber, 1969) and modalities (e.g., Altmann, Dienes, & Goode, 1995; Tunney & Altmann, 2001; cf. Conway & Christiansen, 2006). The present study differs from these previous investigations in two key ways. First, we explore VSL, which is importantly different from AGL, particularly in that units can be parsed only from undifferentiated sequences or layouts via the extraction of statistical regularities. Second, we explore a type of transfer that is fundamentally different: Whereas these AGL studies explored whether learned grammars can be expressed with entirely novel elements, here we explore whether knowledge of statistical regularities can be expressed when the *manner* in which the same elements are presented—especially their local spatiotemporal context—is novel. In particular, we explore generalization in VSL by testing whether learned associations transfer across space and time.

Temporal VSL

The study of temporal VSL emerged from infant studies of auditory statistical learning between syllables in an undifferentiated speech stream (Saffran, Aslin, & Newport, 1996). In a visual variant (Fiser & Aslin, 2002; see also Olson & Chun, 2001), adult observers were presented with an animation in which a single object moved horizontally across the screen, continuously cycling back and forth and changing its shape each time it passed behind a central occluder (see Figure 1c); this animation lasted for only a few minutes, and there was no specific task. The sequence of shapes consisted of temporal *triplets* in which the same three shapes always appeared in the same order (e.g., ABCGHIDEFABC . . .). Critically, only this statistical regularity demarcated the triplets, as the intershape delay was constant. After this passive exposure, observers completed a surprise familiarity test that pitted triplets (e.g., ABC) against foil sequences of three shapes that were individually as familiar as the triplet shapes but had a joint probability of 0 (e.g., AEI). Observers reliably identified the triplets as more familiar than the foil sequences, indicating robust VSL. Such processing requires attention, but can operate without intent or awareness, and despite a secondary task (Turk-Browne et al., 2005).

Spatial VSL

VSL can also extract spatial regularities when multiple objects are viewed simultaneously. In an initial study of this type (Fiser & Aslin, 2001; see also Chun & Jiang, 1998), observers were presented with a series of 3×3 grids, each containing six shapes (see Figure 3b). Unbeknownst to observers, each shape was always spatially paired with one of the other shapes. For example, Shape A might always appear immediately above and to the right of Shape B. However, because three pairs were always spatially interleaved in each grid, the pairs could only be segmented on the basis of their statistically reliable spatial relations. At test, observers were presented with the original spatial pairs pitted against foils consisting of two equally familiar shapes that had appeared together less frequently. Again, observers demonstrated a robust preference for the actual pairs, illustrating spatial VSL.

The Present Study

All previously published studies of VSL, to the best of our knowledge, presented stimuli at test in the same spatiotemporal manner as they were encountered during learning—with learned temporal sequences probed with identical test sequences (in which elements were presented in the same order as during learning), learned spatial layouts probed with identical test layouts, etc. (Note that this is independent of whether VSL can be expressed with *elements* that are modified along other dimensions—e.g., if items are presented at test in new colors; Turk-Browne et al., 2005; Turk-Browne, Isola, et al., 2008.) As a result, it is not clear just what is being learned and, in particular, how far (or even whether) the resulting mental representations can be generalized beyond the spatiotemporal character of the input. Here, we explore this issue in *transfer* experiments, whereby the manner in which the stimuli are presented differs between learning and test. The underlying logic of such studies is that if VSL transfers to new contexts that disrupt some feature of the initial context, then that feature is not an intrinsic part of the learned representation.

Experiment 1: Reversing Temporal Sequences

Previous studies of temporal VSL probed learning with temporal sequences in which the order of elements is preserved between familiarization and test. Here, in contrast, the familiarization displays present single shapes one at a time in reliable sequences, but the order of the shapes in some test displays is reversed. Forward Triplets were presented at test in the same order as during familiarization (e.g., ABC → ABC), but Backward Triplets were presented at test in a reversed order (e.g., ABC → CBA). Following other studies, we used forced-choice familiarity judgments to assess statistical learning. During each test trial, observers viewed two subsequences—a temporal triplet and a foil containing familiar shapes that never appeared in sequence—and reported which subsequence seemed more familiar.

Because observers were equally familiar with all of the individual shapes, any familiarity preference for the Forward Triplets during test demonstrates learning of the joint probabilities between shapes during familiarization, as in previous studies.[2] The Backward Triplets then provide a critical test of how flexible this learning was with respect to temporal order. Consider three possible outcomes: (a) Failure to express any preference for the

---

[1] At a recent conference presentation, Conway, Goldstone, and Christiansen (2007), however, did observe a marked degree of specificity—and a lack of any generalization—purely within the spatial domain. Thus, the available data on three forms of associative visual learning (statistical learning, contextual cueing, and perceptual learning) all suggest that visual learning results in representations that are highly specific to the learning context.

[2] Following other studies, we term this *statistical learning*, because the test sequences could only be discriminated on the basis of the relationships between shapes (i.e., by their nonzero joint probabilities) and not on the basis of shapes themselves. Note that although this is perhaps the most direct test of statistical learning, researchers have also tested for the learning of even subtler statistics, such as relative joint probability and conditional probability.

Backward Triplets would provide evidence for highly specific representations in which temporal order is a defining feature; (b) a weaker preference for the Backward Triplets versus foils than for Forward Triplets versus foils would provide evidence that the learned representations did include temporal information but could nevertheless be generalized to different orders; and (c) equivalent results for Forward and Backward Triplets would be consistent with the possibility that the learned representations could be completely generalized from the precise temporal context.

Method

Participants

Eight naive undergraduate students participated for course credit.

Apparatus and Stimuli

We used custom software written with the VisionShell graphics libraries (Comtois, 2006) to present stimuli. Observers without head restraints sat approximately 60 cm from a computer monitor. Displays were constructed from a set of 12 black shapes that researchers have used previously in studies of VSL (e.g., Fiser & Aslin, 2001), with each shape subtending 3.0°. Each sequence began with the object located in the center of the display, where it immediately began moving to either the left or to the right at 10.3°/s. When the center of the object was 11.5° 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 green occluder, subtending 8.6° × 5.1°, 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.

Procedure

Learning phase. Each of the 12 shapes was assigned to a unique position in one of four *triplets* (e.g., ABC), sequences of three shapes that always appeared in the same order (Figure 1a). The full sequence of shapes was generated by randomly interleaving 24 repetitions of each triplet with two constraints: (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 was .083, and the joint probability of a sequence of three shapes spanning triplets (e.g., CGH) was .027 (see Figure 1b). Observers watched a 5-min animation of 288 shapes, presented one at a time (for 1 s each) in the manner described above (see Figure 1c).

Test phase. Observers then completed a two-alternative forced-choice familiarity test, pitting triplets against foil groups of three familiar shapes that had never appeared sequentially during familiarization (e.g., ABC vs. AEI). Half of the triplets presented during the test phase appeared in the same order as during familiarization (Forward Triplets; e.g., ABC), whereas the other half were presented in reversed order (Backward Triplets; e.g., CBA; Figure 2a). To equate the frequency of alternatives across test trials, we tested the Backward Triplets against foils in which the elements were also reversed (e.g., IEA). The four Forward Triplets were each tested once against each of the four Forward Foils, and the four Backward Triplets were each tested once against each of the four Backward Foils, for a total of 32 test trials. The shapes in the triplet and foil sequences were presented one at a time with the same timing as during familiarization; the two-alternative three-item sequences during each test trial (i.e., the triplet and the foil) were separated by a 1-s pause. Whether the triplet appeared first or second was randomized across trials. Observers pressed one key if the first alternative seemed more familiar and a different key if the second alternative seemed more familiar.

Results and Discussion

Our measure of VSL was the percentage of test trials in which the triplet was chosen as more familiar than the foil (chance = 50%). As can be seen in Figure 2b, observers preferred triplets over foils for both Forward Triplets, 70%, $t(7) = 3.00$, $p = .02$, $d = 1.06$, and Backward Triplets, 70%, $t(7) = 2.78$, $p = .03$, $d = 0.98$, and these levels of learning did not differ (and were in fact identical). Of course, these results do not suggest that temporal information is completely absent from such representations, and the next experiment demonstrates that it can still be accessed in other contexts. What these results do demonstrate, however, is that the representations that result from VSL of temporal sequences of objects *can* be applied in an order-invariant manner.

Experiment 2: Reversing Temporal Sequences When Order Information Is Required

How flexible is the generalization observed in Experiment 1? In particular, do these results indicate that order information is entirely absent from the learned representations, or might it be the case that learning generalized to a new order in Experiment 1 because order was not necessary to distinguish triplets from foils? Here we ask whether observers can access temporal order information in the same learning context when order information is required to distinguish triplets from foils at test. We implemented this in the most direct possible way, by directly pitting Forward Triplets (e.g., ABC) against Backward Triplets (e.g., FED).

Method

Fourteen new undergraduate students participated for course credit. The apparatus, stimuli, and familiarization procedure were identical to those of Experiment 1. The test phase procedure was also similar to that of Experiment 1, except that each trial tested a Forward Triplet against a Backward Triplet (e.g., ABC vs. FED). Each Forward Triplet was tested against each Backward Triplet twice, for a total of 32 test trials.

Results and Discussion

Because observers could only distinguish the test alternatives on the basis of the order in which the shapes within a triplet had been presented during familiarization, any preference for the Forward Triplets would demonstrate that the learned representations of triplets contained at least some temporal order information, despite the fact that such knowledge was not revealed in Experiment 1. Observers showed a preference for Forward Triplets over Backward Triplets (Figure 2b) by choosing the Forward alternative on significantly more than 50% of the test trials, 61%, $t(13) = 2.42$,

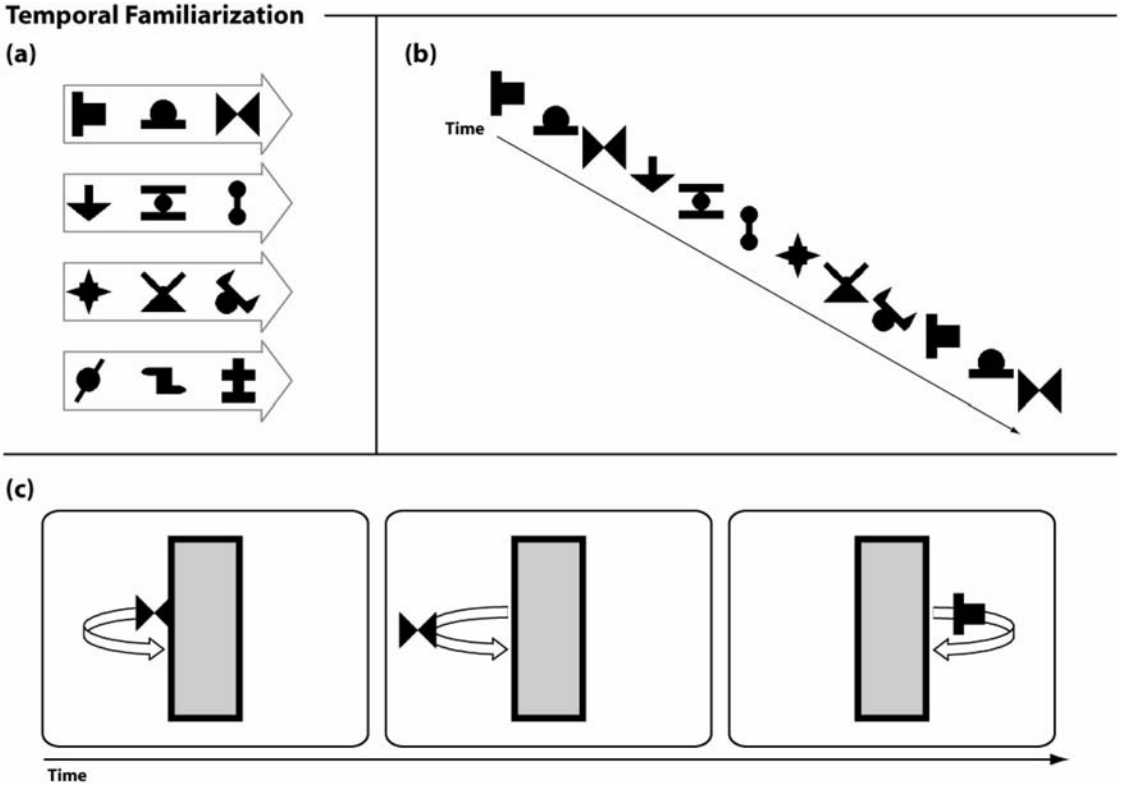

Figure 1. Temporal familiarization of Experiments 1–4 is depicted. The 12 shapes are shown arranged in four ordered triplet sequences of three shapes each (a). The stream consisted of shapes presented one at a time, arranged into triplets but without any other segmentation cues (b). During familiarization in Experiments 1 and 2, shapes were presented one at a time in a disoccluding–occluding manner from the center of the display (c). In Experiments 3 and 4, the shapes simply appeared one at a time in the center of the display (not shown).

$p = .03$, $d = 0.65$. In conjunction with Experiment 1, these results demonstrate that the degree of generalization in VSL (at least in the case of temporal order) is a flexible function of the demands of the test context. When learning can be expressed without consideration of temporal order, it generalizes to new orders (as in Experiment 1), but when learning can be expressed only by considering temporal order, temporal order information is used (as in this experiment).

Experiment 3: From Time to Space

The robust transfer across temporal order in Experiment 1 suggests that VSL may allow for a considerable degree of generalization across spatiotemporal details. Does VSL of temporal sequences even require sequential tests? Previous studies of temporal VSL probed learning with purely temporal displays, in which shapes were presented one at a time during both familiarization and test. Here, in contrast, the familiarization displays present single shapes one at a time, but the test displays present all shapes simultaneously.

Method

Eight new undergraduate students participated for course credit. The apparatus, stimuli, and familiarization procedures were identical to Experiments 1 and 2, with one exception. To eliminate any spatial information during familiarization, we presented individual shapes for 1 s each in the center of the screen, without the occluders or the movement that we used in Experiments 1 and 2. In the test phase, observers completed a two-alternative forced-choice familiarity test, pitting triplets against foil groups of three familiar shapes that had never appeared sequentially during familiarization (e.g., ABC vs. AEI). Critically, the three shapes in each alternative were presented simultaneously in a static triangular configuration that subtended 8.1° × 7.5° (Figure 2c), with the positions in the triangle randomized. The triplet and foil displays were presented for 3 s each, separated by a 1-s pause, and their order was randomized across trials. Observers pressed one key if the first configuration seemed more familiar and a different key if the second configuration seemed more familiar. Every triplet was tested against every foil twice (for a total of 32 test trials), equating frequency of the alternatives during test.

Results and Discussion

Because observers were equally familiar with all of the individual shapes, any familiarity preference for the spatial triplets during test (after temporal learning) demonstrates learning of shape triplets from the initial temporal sequence. Observers showed robust

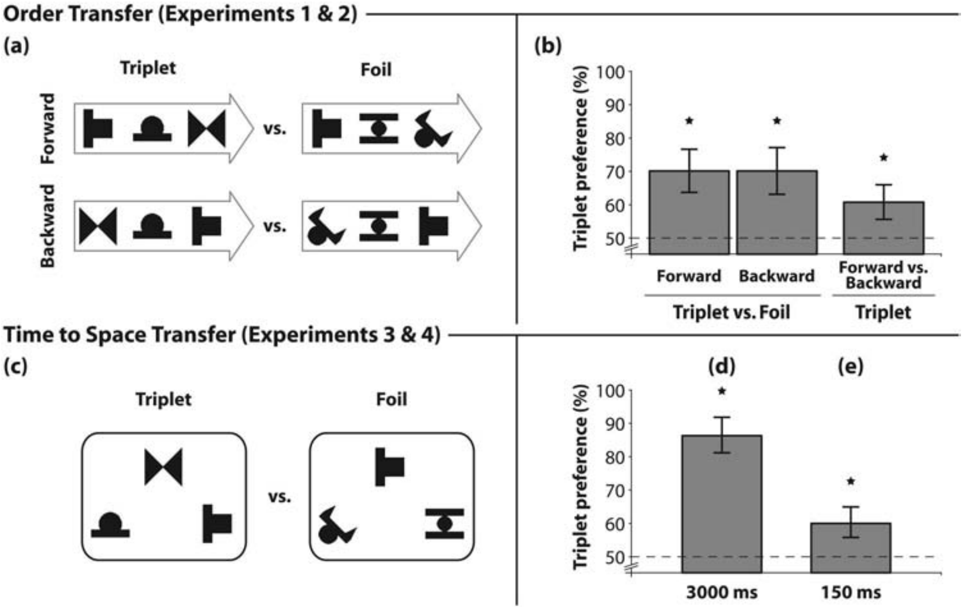

Figure 2. Temporal transfer tests in Experiments 1–4 are shown. In the test phase of Experiment 1 (a), each trial contained a triplet presented in the same order as during familiarization (Forward) or in reversed order (Backward). Both were tested against a foil that was repeated an equal number of times during the test phase. In the test phase of Experiment 2 (b), observers demonstrated a reliable preference for the Forward triplets compared to Backward triplets. In the test phase of Experiments 3 and 4 (c), triplets and foils were presented simultaneously in a static spatial configuration. Observers demonstrated a reliable preference for spatial triplets relative to foils when the spatial configurations at test were presented for 3,000 ms in Experiment 3 (d) and for 150 ms in Experiment 4 (e). The dashed line indicates chance performance; error bars reflect one standard error of the mean. $^{*}p < .05$.

statistical learning (Figure 2d) that differed from chance, 86%, $t(7) = 7.52$, $p < .01$, $d = 2.66$. This demonstrates that VSL of temporal sequences of objects results in flexible mental representations that can later be applied to static spatial layouts.

Experiment 4: From Time to Space—*Quickly*

A potential problem with Experiment 3 is that temporal information may have existed during the test phase due to sequential eye movements during the long (3-s) presentation times. Although all three shapes from a triplet were presented simultaneously, observers may have scanned them sequentially. Here we eliminate this possibility by presenting the test displays more briefly than in any previous studies and, in particular, too briefly for sequential eye movements or shifts of attention.

Method

Fourteen new undergraduate students participated for course credit. All stimuli and procedures were identical to Experiment 3 except that test displays were each presented for 150 ms and observers were instructed to fixate on a central cross that appeared throughout each test trial.

Results and Discussion

As in Experiment 3, observers showed robust expression of VSL (Figure 2e) by choosing the triplet over the foil more often than would be expected by chance, 60%, $t(13) = 2.39$, $p = .03$, $d = 0.64$. Unsurprisingly, performance was lower in Experiment 4 than in Experiment 3, $t(20) = 3.99$, $p < .01$, $d = 1.80$, presumably because the quick presentations reduced observers' ability to resolve shapes in the periphery (a difficulty that was noted by many of the observers). The fact that performance was still above chance, in any case, rules out the possibility that the transfer of VSL from time to space observed in Experiment 3 could have resulted solely from haphazard patterns of eye movements. This result provides even stronger evidence that foveal sequential VSL can later be expressed in static displays of peripheral shapes.

Experiment 5: From Space to Time

Whereas Experiments 3 and 4 tested for transfer from time to space, here we tested for transfer from space to time. All previous studies of spatial VSL have probed learning with spatial displays, in which shapes were presented simultaneously. Here, in contrast, the familiarization displays presented all objects simultaneously, but the test displays presented single shapes one at a time. In addition, we moved away from familiarity judgments at test and instead used an implicit response time measure (based on Turk-Browne et al., 2005; cf. Olson & Chun, 2001). During each test trial in that study, observers were shown a single stream of shapes, and they had to press a key when they detected a prespecified target shape. Successful VSL was observed in this design in terms of speeded responses to target shapes when they were preceded by shapes that had also preceded the target during familiarization.

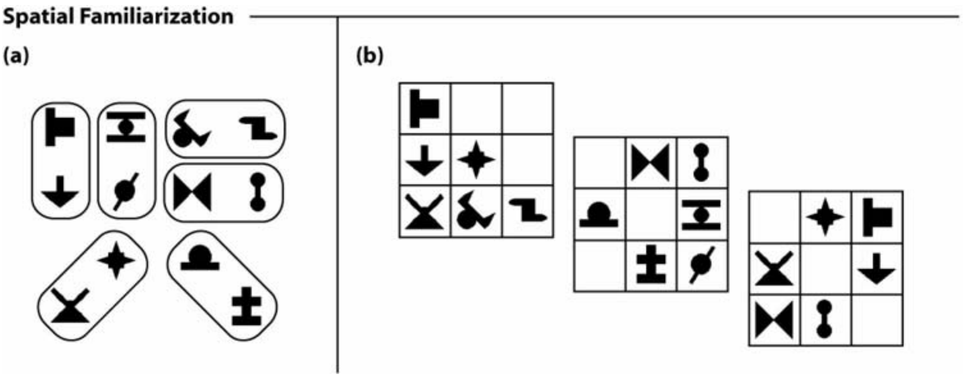

Figure 3. Spatial familiarization in Experiment 5 is depicted. The 12 shapes are arranged into six spatial pairs (a). Each grid during familiarization (b) consisted of three interleaved pairs (one of each orientation).

Here we employed this method to test whether spatial pairs from familiarization also give rise to a temporal cueing benefit at test.[3]

Method

Eight new undergraduate students participated for course credit. The apparatus and shape stimuli were identical to the other experiments, except that stimuli were presented with the Psychophysics Toolbox for MATLAB (Brainard, 1997; Pelli, 1997). During learning, shapes were presented in a 3 × 3 grid that subtended 14.0°.

Procedure

Learning phase. Each of the 12 shapes was assigned to one of six spatial pairs in either a horizontal, vertical, or diagonal orientation (Figure 3a). Observers watched a 7-min series of 144 grids, each presented for 2 s and separated by a 1-s pause (Figure 3b). The grids were constructed by choosing one pair of each orientation, placed such that each pair was adjacent to at least one other pair. These constraints limited each pair to four possible locations on the grid, yielding 144 distinct grids. Because each pair appeared in half of the displays, the joint probability of the two shapes within a pair was 0.500. The joint probabilities of unpaired adjacent shapes (i.e., one shape from each of two spatial pairs) were variable, but much lower (typically less than .02, as calculated by Fiser & Aslin, 2001). Because we tested for temporal effects, the order in which the grids appeared was constrained such that no paired shapes could appear in the same grid cell on two subsequent grids.

Test phase. After the learning phase, observers completed an implicit response time test (Figure 4a). At the beginning of each test trial, observers were presented with 1 of the 12 shapes from the learning phase that served as the target for that trial. Upon pressing a key, observers were presented with a stream of shapes appearing one at a time in the center of the screen for 400 ms followed by a 400-ms pause.

Each stream consisted of a total of 14 shapes, including 12 novel shapes that had not been viewed during learning (to prevent reinforcing or weakening learned relationships during test) and a *critical pair*—a sequence of 2 shapes, the second of which was the target. The target appeared randomly in Positions 3–13 in the stream. The first shape in the critical pair was either the shape with which the target had been spatially paired during learning or a shape from the other pair of the same orientation, which had never appeared in the same grid. Each of the 2 shapes in a spatial pair served as both a cue and a target for the other shape on different trials. Crucially, the foil pairings were also reversed to equate joint probability during test. Each of the 24 possible cue–target pairings was presented twice for a total of 48 trials.

Observers responded as quickly and accurately as possible by pressing a key whenever they detected the target shape. Responses that occurred before the critical pair or three or more shapes after the target were considered false alarms. In four preliminary practice trials, observers searched for one of the novel shapes.

Results and Discussion

Target detection performance was excellent, with a total of only 1.1 false alarms on average (range: 0–5). Response times were analyzed as a function of the preceding shape. As shown in Figure 4b, the target was detected reliably faster when it was preceded by the shape with which it had been spatially paired during familiarization (350 ms) compared with an equally familiar shape with which it had never appeared (380 ms), $t(7) = 2.45$, $p = .04$, $d = 0.87$. This difference was not the result of a speed–accuracy tradeoff, as slightly fewer errors were made in the faster condition (0.5 vs. 0.6). This demonstrates that VSL of spatial layouts results in flexible mental representations that can later be applied to temporal sequences.

General Discussion

The experiments reported here constitute an initial demonstration of how the representations generated by VSL can be generalized from the spatiotemporal context of the initial displays. Indeed, the successful transfer across temporal order and between space and time observed in our experiments highlights the importance of such manip-

---

[3] We deem this type of implicit test to be an improvement over the standard familiarity judgments used by researchers in the previous experiments and in previous studies of VSL (following Fiser & Aslin, 2001, 2002). Familiarity judgments may be sensitive to the results of implicit learning, but they can also be readily influenced by explicit strategies and conscious guesswork. In contrast, the implicit test developed in this experiment can be used by researchers to assess the results of temporal VSL in a way that is more insulated from explicit strategies. This paradigm obviously cannot apply to the spatial tests as used in Experiments 3 and 4, however.

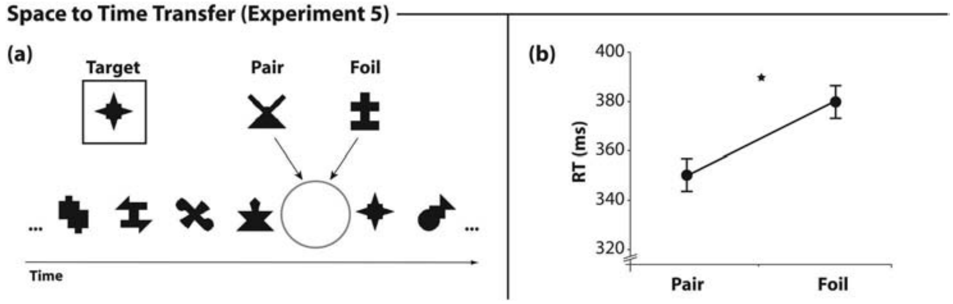

Figure 4. Spatial transfer test in Experiment 5 is shown. On each test trial, observers memorized a particular target, to which they then made a speeded response when it appeared in a stream of shapes presented rapidly, one at a time (a). Critically, the target was either preceded by the shape with which it was spatially paired during learning (in the pair condition), or with an equally familiar but spatially unpaired shape (in the foil condition). Targets were detected faster in the pair condition (b). Error bars reflect one within-subject standard error of the mean difference. $^{*}p < .05$.

ulations, especially because learning was expressed in some contexts (Experiments 3–5) that bore little resemblance to what was initially encountered. Moreover, the implicit test that we employed in Experiment 5 demonstrates that generalization is not isolated to explicit comparisons, as has sometimes been claimed (e.g., Conway & Christiansen, 2006; cf. Maddox & Ashby, 2004).

One can perhaps understand how generalization might be natural when the dimension being abstracted over is highly variable during learning (as in the transfer to new syllables during AGL; e.g., Marcus, Vijayan, Rao, & Vishton, 1999). In this sense, however, the transfer observed here is all the more striking given that the temporal or spatial information during familiarization was always present and indeed was essential to defining the initial statistical regularities.

The fact that VSL transferred across space and time in this study implies that VSL has a purely associative component, which in turn suggests that it may have previously unforeseen connections to the rich body of research on associative learning. These other literatures have also suggested ways in which spatial and temporal learning may reflect the same types of underlying processing. Although we typically think of conditioning experiments as inherently sequential, in fact simultaneous exposures can yield similar (and sometimes stronger) results (e.g., Rescorla, 1980). This possibility is also suggested in demonstrations of associative symmetry in paired-associate learning (Asch & Ebenholtz, 1962; Kahana, 2002) and is consistent with demonstrations in humans that prediction is sometimes no better than “retrodiction” in explicit judgments (Jones & Pashler, 2007). However, the generalization away from time observed here may not hold in other learning contexts. For example, when a conditioned stimulus follows an unconditioned stimulus in a backward conditioning procedure, the conditioned stimulus fails to elicit the conditioned response (and may actually inhibit its expression; e.g., Moscovitch & LoLordo, 1968). These examples suggest that it will be particularly important for future work to relate the study of VSL to other forms of associative learning, because both their similarities and their differences could be informative and provide a useful synthesis: Whereas previous studies of implicit learning often explicitly resonate with older traditions of associative learning, this older literature is almost never discussed or even cited in the current incarnation of VSL research (cf. Perruchet & Pacton, 2006). In fact, recent neuroimaging evidence suggests that the processes underlying VSL may overlap with those of other forms of associative and implicit learning (Turk-Browne, Johnson, et al., 2008).

Of course, our results do not preclude the possibility that VSL also generates more specific mental representations (cf. Chun & Jiang, 1999) or that repeating details of the initial spatiotemporal context could facilitate the resulting expression of learning. Because our goal in this project was simply to assess whether there exists generalization across spatiotemporal details in VSL, we did not directly compare learning with and without changes to the spatiotemporal context between familiarization and test, but future studies could do so. The resulting tests, however, would have to be between-subjects because two distinct test phases would be required. One such test reported here was in Experiment 1, wherein both Forward and Backward Triplets were intermixed during test, resulting in identical expressions of learning. However, we also observed that temporal order information was still available when it was required, as in the direct test of Forward versus Backward Triplets in Experiment 2. This indicates that the degree of generalization that will be observed in such contexts is itself flexible and can be modulated by task demands. Future research could thus explore the factors that induce or inhibit generalization. For example, transfer may also depend on the type of statistics involved and therefore may or may not occur in the context of conditional probabilities.

On the basis of our study, however, it will be important to search for such specificity beyond the suggestions that learning may not always transfer between modalities (Conway & Christiansen, 2006).[4] For example, statistical learning has been proposed as a mechanism for initial word segmentation during language acquisition, but temporal order is clearly a fundamental property of language. Thus, it will be important to determine whether temporal order information is used in such contexts, because any inability to

---

[4] Note that our results do not in any way conflict with other demonstrations that the representations resulting from statistical learning are modality- or stimulus-specific (Conway & Christiansen, 2006). The representations of the statistically segmented visual subsequences (i.e., the visual “words”) could be modality-specific (or specific to the particular elements that we used) and still generalized from the manner in which those elements are presented (i.e., their spatiotemporal context in the present study).

apply such knowledge in linguistic contexts would render statistical learning ill-suited for word segmentation.

At the same time, the remarkable degree of generalization observed here may be important because real-world visual environments are constantly in flux. Consider the sequence of people you might see each day during a train commute. This sequence could be highly regular, but countless other aspects of the situation may be changing (e.g., their clothes, where they sit in the car, the presence of other haphazard people and objects, the view out the window). As a result, learning that could only be expressed during later exposure to identical input patterns would never reach such triggers. Thus, generalizing across some features of the input when constructing mental representations of statistical regularities may help VSL operate in natural visual environments. And as the present results indicate, the scope of this generalization may be considerable.

References

Altmann, G. T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899–912.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society, 106,* 135–163.

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

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106,* 1382–1407.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10,* 360–365.

Comtois, R. (2006). VisionShell PPC [Software libraries]. Cambridge, MA: Author.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17,* 905–912.

Conway, C. M., Goldstone, R. L., & Christiansen, M. H. (2007). Spatial constraints on visual statistical learning of multi-element scenes. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (pp. 185–190). Nashville, TN: Cognitive Science Society.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology, 78,* 2889–2894.

Fahle, M., & Poggio, T. (Eds.). (2002). *Perceptual learning.* Cambridge, MA: MIT Press.

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

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

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

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.

Jiang, Y., & Song, J.-H. (2005). Hyperspecificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1439–1448.

Jones, J., & Pashler, H. (2007). Is the mind inherently forward-looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review, 14,* 295–300.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30,* 823–840.

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.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes, 66,* 309–332.

Marcus, G., Vijayan, S., Rao, B., & Vishton, P. (1999). Rule learning by 7-month-old infants. *Science, 283,* 77–80.

Marr, D. (1982). *Vision.* San Francisco: WH Freeman.

Moscovitch, A., & LoLordo, V. M. (1968). Role of safety in Pavlovian backward fear conditioning procedure. *Journal of Comparative and Physiological Psychology, 66,* 673–678.

Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1299–1313.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

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

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81,* 115–119.

Rescorla, R. A. (1980). Simultaneous and successive associations in sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 6,* 207–216.

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

Tunney, R., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 614–639.

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., Johnson, M. K., Chun, M. M., & Scholl, B. J. (2008). *Neural evidence of incidental statistical learning.* 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.

von Helmholtz, H. (1925). Treatise on physiological optics. In J. P. Southall (Ed. & Trans.), *Treatise on physiological optics Vol. III.* New York: Dover. (Original work published 1910)

Received August 17, 2007
Revision received December 7, 2007
Accepted February 10, 2008 ■