Journal of Experimental Psychology: General
2005, Vol. 134, No. 4, 552–564

0096-3445/05/$12.00 DOI: 10.1037/0096-3445.134.4.552

The Automaticity of Visual Statistical Learning

Nicholas B. Turk-Browne, Justin A. Jungé, and Brian J. Scholl
Yale University

The visual environment contains massive amounts of information involving the relations between objects in space and time, and recent studies of visual statistical learning (VSL) have suggested that this information can be automatically extracted by the visual system. The experiments reported in this article explore the automaticity of VSL in several ways, using both explicit familiarity and implicit response-time measures. The results demonstrate that (a) the input to VSL is gated by selective attention, (b) VSL is nevertheless an implicit process because it operates during a cover task and without awareness of the underlying statistical patterns, and (c) VSL constructs abstracted representations that are then invariant to changes in extraneous surface features. These results fuel the conclusion that VSL both is and is not automatic: It requires attention to select the relevant population of stimuli, but the resulting learning then occurs without intent or awareness.

Keywords: statistical learning, implicit learning, selective attention, nonadjacent dependencies, specificity

The goal of visual processing is to recover information about the structure of the local environment, and what makes this problem difficult is that there is both too little and too much incoming information. The available information is rarely sufficient to definitively infer the structure of the world without making several heuristic assumptions, but at the same time we would be paralyzed if we were perceptually confronted with the massive amount of available sensory input. Thus, visual perception must be selective, and we are consciously aware of only a small part of the visual world at any time. Because the bulk of visual processing occurs beneath the level of awareness, we—as both laypeople and vision scientists—are often surprised at the scope and subtlety of such processing. In this article we explore the nature of a particularly powerful form of unconscious visual processing, concerning the relationships between objects over time.

Statistical Learning

Associative learning of the relationships between objects has played a prominent role in several areas of psychology, but has been less widely explored in the context of human perception and cognition. Interest in this type of learning has flourished in recent years, however, due to the discovery of a novel form of statistical learning.

The initial report of statistical learning demonstrated that 8-month-old infants can learn subtle statistical relationships that hold across a sequence of nonsense syllables in a pseudospeech stream (Saffran, Aslin, & Newport, 1996). On the surface, this stream appeared to contain a random sequence of syllables (e.g., "pi-go-la-bi-ku-ti- . . . ," in which each syllable is denoted by a capital letter below), but in fact the sequence contained statistical regularities. The 12 possible syllables were divided into triplets (i.e., *ABC*, *DEF*, *GHI*, *JKL*), such that the order of the triplets was randomized, but each triplet itself always progressed through the same three syllables in the same order (e.g., *ABC, GHI, DEF, ABC, JKL* . . .). After only 2 min of exposure to such a sequence, infants were able to discriminate the triplets they had heard most often (e.g., *GHI*) from both (a) sequences they had never encountered (e.g., *AEI*) and (b) other "accidental" three-syllable sequences they had heard less often (e.g., *BCG*, from when triplet *GHI* happened to follow triplet *ABC*). In statistical terms, this result requires the infants to have encoded the greater joint probability of the triplets during familiarization. (Later experiments demonstrated that infants in such situations could also learn more complex conditional probabilities when joint probabilities were equated; Aslin, Saffran, & Newport, 1998).

Auditory statistical learning of this type excited many researchers because of its potential to shed light on aspects of language learning, such as how young children learn to segment words from a continuous speech stream (e.g., Bates & Elman, 1996; Seidenberg, 1997). The possibility that statistical learning is involved in such linguistic processing has been supported by (a) additional demonstrations that it can operate even over nonadjacent dependencies (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004); (b) suggestions that statistical computations in pseudospeech streams are in some cases triggered by the lack of acoustic segmentation cues (Peña, Bonatti, Nespor, & Mehler, 2002; cf. Perruchet, Tyler, Galland, & Peereman, 2004); (c) the observation that

---

Nicholas B. Turk-Browne, Justin A. Jungé, and Brian J. Scholl, Department of Psychology, Yale University.

For helpful conversation and comments on drafts of this article, we thank Dick Aslin, Luca Bonatti, Tim Brady, Jen Catena, Marvin Chun, Koleen McCrink, Steve Mitroff, and Jenny Saffran. Nicholas B. Turk-Browne was supported by a foreign Natural Sciences and Engineering Research Council of Canada Post-Graduate Scholarship, and Brian. J. Scholl was supported by National Science Foundation Grant BCS-0132444.

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

statistical regularities that are characteristic of the world's natural languages are easier to learn than those that are not (Saffran & Thiessen, 2003); and (d) suggestions that statistical information that is particularly relevant to word segmentation (e.g., involving consonants, in contrast to vowels) is more easily learned (Bonatti, Peña, Nespor, & Mehler, 2005). Other evidence, however, suggests that this ability may not be tied to language, per se. For example, statistical learning also operates over nonlinguistic auditory stimuli such as musical tones (Creel et al., 2004; Saffran, Johnson, Aslin, & Newport, 1999) and in nonhuman primates (Hauser, Newport, & Aslin, 2001; Newport, Hauser, Spaepen, & Aslin, 2004). Moreover, recent computational modeling suggests that segmentation via this type of statistical learning may not scale up to the speech signal involved in natural language acquisition (Yang, 2004).

Visual Statistical Learning

Another hint that the processes underlying statistical learning are not unique to language comes from analogous studies of visual statistical learning (VSL), in which observers—both adults and infants—are able to implicitly learn subtle statistical relationships among visual objects in both space (Chun & Jiang, 1998; Fiser & Aslin, 2001, 2002b) and time (Fiser & Aslin, 2002a; Fiser, Scholl, & Aslin, 2004; Kirkham, Slemmer, & Johnson, 2002; Olson & Chun, 2001).

One of the clearest demonstrations of VSL in adults—which was designed as a visual analog of the original studies with infants, and which serves as the foundation for the current experiments—involved statistical learning of the temporal relationships among sequentially presented shapes (Fiser & Aslin, 2002a). Observers viewed an animation in which a single object moved on a horizontal path across the screen, continuously cycling back and forth behind a central occluder and changing its shape each time it passed behind the occluder. Participants watched this animation for 6 min, with no specific task. In fact, unbeknownst to the participants, the sequence of shapes was not fully random, but rather was structured as sequences of triplets (canonical three-shape subsequences) using the same design as the auditory statistical learning study of Saffran et al. (1996) described above (e.g., *ABC*, *GHI*, *DEF*, *ABC*, *JKL*). To determine whether observers were sensitive to this statistical structure, the experimenters had them complete a surprise two-interval forced-choice (2IFC) familiarity task that pitted triplets (e.g., *ABC*) against foil sequences consisting of a string of three shapes that had joint probabilities equal to 0 (e.g., *AEI*). Observers correctly identified 95% of the triplets as more familiar, indicating robust statistical learning of visual temporal sequences.

Attention and Automaticity in Statistical Learning

One of the most exciting aspects of statistical learning is the possibility that it may reflect an automatic underlying perceptual process, rather than a higher-level intentional learning strategy. Although the automaticity of statistical learning has received little direct study, several aspects of the experimental designs described above are relevant to this issue. In particular, note that participants in these experiments were typically familiarized with the stimulus sequences without any orienting instructions beyond simple requests to watch (or listen to) the displays. Moreover, it has frequently been reported that participants were completely unaware of the underlying statistical structure in the sequences, despite the fact that it fueled their familiarity judgments (e.g., Fiser & Aslin, 2002a) or speeded their later visual performance in search tasks (e.g., Olson & Chun, 2001). As a result of such factors, this type of statistical learning has been thought to proceed "automatically" (Fiser & Aslin, 2002a, p. 458), "incidentally" (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997, p. 101), "spontaneously" (Fiser & Aslin, 2001, p. 502), and "as a byproduct of mere exposure" (Saffran et al., 1999, p. 30).

On the other hand, there are several reasons to suspect that statistical learning may not be entirely automatic. In the first place, the lack of any overt task during the familiarization phase may not have prevented observers from intentionally processing the structure of the displays. (Indeed, participants in statistical learning studies are often initially uncomfortable with this undirected viewing and seem to think, "How can I do nothing?") In addition, because there is no task involved during the familiarization, the experimental setup allows or even encourages observers to attend to each of the familiarization shapes. This seems important, because a major theme in visual cognition over the last decade has been that attention mediates many different types of perceptual processing and indeed is often necessary for becoming consciously aware of an object in the first place (e.g., Mack & Rock, 1998; Most, Scholl, Clifford, & Simons, 2005). Thus it remains unclear from these initial experiments whether this "automatic" statistical learning requires that participants attend carefully to the stimulus sequences, or whether this process operates in a pre-attentive fashion more characteristic of lower-level visual processes.

To our knowledge, only two results in the literature bear on the role of attention in statistical learning of this type. First, both children and adults are able to learn joint probabilities in "background" auditory streams even while performing a primary visual task (Saffran et al., 1997). The motivation for this early study was to see whether statistical learning would occur even when participants were not oriented to the fact that the auditory stream was relevant to the experiment at all, and this study did not directly manipulate or test attention. Nevertheless, it seems plausible that participants were less attentive to the background auditory stream in this condition, and in fact this study has been interpreted as evidence that statistical learning does not require attention. Second, a recent study of spatial statistical learning demonstrated that certain pairwise-associations between target and distractor shapes were learned only when observers were unaware of where the target would appear (Baker, Olson, & Behrmann, 2004). When the target's location was known in advance (and thus was presumably spatially attended, perhaps even fixated), statistical associations involving other shapes that were most likely outside the focus of attention were not learned.

The Influence of Attention: Resource Versus Selection

What might explain the contrast between the results of these two studies? A related debate about the role of attention in other types of implicit learning may provide a useful distinction that can reconcile the fact that attention was perhaps unnecessary in the studies of Saffran et al. (1997) but was important in the studies of Baker et al. (2004). On the topic of implicit learning of perceptual motor sequences and target–distractor pairings, initial studies demonstrated that a concurrent task disrupted performance in a serial reaction time paradigm, suggesting that attention was required for learning (Nissen & Bullemer, 1987). Several subsequent studies questioned this conclusion, however, and demonstrated that in other circumstances, dual tasks do not interfere with implicit sequence learning (e.g., Cohen, Ivry, & Keele, 1990; Frensch, Lin, & Buchner, 1998; Stadler, 1995).

These conflicting demonstrations were then reconciled in an experiment that exploited the distinction between attention as a resource versus attention as a mechanism of selection (Jiménez & Méndez, 1999). This distinction highlights the difference between two of the most fundamental aspects of attention: that it can be applied to some stimuli and not others, and that this application has a limited capacity (e.g., Johnston & Dark, 1986; Pashler, 1998). Jiménez and Méndez (1999) experimentally demonstrated that capacity limitation per se did not affect implicit learning in dual-task situations, but that selective attention to some features but not others resulted in implicit learning of only the attended features. This may also explain why attention is able to modulate implicit learning in spatial contextual cueing tasks, in which repeatedly encountered spatial layouts of distractors are predictive of target location and thus speed visual searches for those targets (Chun & Jiang, 1998, 2003). When observers selectively attended to distractors of only one of two colors, for example, then only distractor layouts of that color facilitated visual search, despite the fact that layouts of both colors were predictive (Jiang & Chun, 2001; Jiang & Leung, 2005).

This distinction between two aspects of attention may explain the contrast between earlier statistical learning studies. In particular, Saffran et al. (1997) may have observed statistical learning despite the lack of explicit orientation to the stream that contained the statistical regularities precisely because their task did not require attentional selection between competing stimuli. (Indeed, their task was not intended to manipulate selective attention). In contrast, the manipulation of attention in Baker et al. (2004) was selective in the sense that observers were biased to specific spatial locations and not others.

The Current Experiments

In the current experiments we manipulate attention in a new way that has more direct implications for the automaticity of VSL. Based on the previous discussion, this goal clearly requires the manipulation of selective attention, but spatial manipulations (as used in Baker et al., 2004) are not ideal for this purpose. The strongest test of automaticity requires that the incoming stimulation from attended and unattended items be equivalent, but spatial manipulations allow for two types of asymmetries in this regard. First, knowledge about the location of a relevant stimulus may simply lead observers to fixate that location, leaving other ("unattended") distractor locations in the periphery, where their structure will not be perceived with the same acuity. Second, a similar asymmetry may arise in spatial displays even if targets and distractors are always presented at equal retinal eccentricities during steady fixation, because there are also resulting differences in attentional resolution (He, Cavanagh, & Intriligator, 1997). The appearance of objects can change depending on whether additional stimuli present in the display are being attended, even controlling for fixation (e.g., Carrasco, Ling, & Read, 2004). Both of these asymmetries may have applied in the study of Baker et al. (2004), and they may directly account for the apparent lack of VSL for unattended stimuli in any such spatial study. In our experiments, we thus used a fundamentally temporal manipulation such that both attended and unattended stimuli were always presented in isolation. In addition, we sought to employ a design that was as similar as possible to previous studies of statistical learning.[1]

The basis for the current experiments was the temporal VSL study of Fiser and Aslin (2002a), but we simplified the stimuli such that the temporal sequences appeared as static shapes presented one at a time in the center of the display. Observers viewed a sequence of geometric shapes, appearing at stimulus onset asynchronies (SOAs) of either 400 or 1,000 ms per shape, depending on the experiment. Half of the shapes were red, and half were green, with a separate pool of shapes for each color (see Figure 1).[2] The sequence of shapes was constructed by randomly intermixing a stream of red shapes with a stream of green shapes (see Figure 2). Unbeknownst to observers, the color streams were constructed from four possible triplets, just as in the studies of Saffran et al. (1996) and Fiser and Aslin (2002a). These triplets comprised the temporal statistical regularities to be learned.

Attention was manipulated by having observers detect shape repetitions in only one of the colors. In a surprise forced-choice familiarity test, triplets from both color streams were then pitted against foil sequences of three shapes from the same color that had never appeared in succession. The test items, both triplets and foils, were presented in all black (Experiments 1A, 1B, and 3), in the same colors as encountered during familiarization (Experiment 2A), or in the opposite colors from those encountered during familiarization (Experiment 2B). In all experiments, our observers had seen each shape an equal number of times; thus the patterns to be learned all involved statistical relationships between shapes, realized in the triplet structure. If VSL is automatic in a strong sense and is truly a product of mere exposure, then observers should be able to pick out triplets of both colors equally well. In contrast, if selective attention is required to gate VSL, then observers may learn only the statistical regularities that involve shapes of the attended color. In either case, the results of this study should help reveal something about the underlying nature of this perceptual process.

Experiment 1A: Manipulating Attention

Visual statistical learning has typically been described as an automatic and implicit process, yet attention has proven critical for

---

[1] In typical studies of statistical learning (e.g., Fiser & Aslin, 2002a; Saffran et al., 1996), and in the studies reported here, the units over which statistical patterns apply are always presented independently. In contrast, the stimuli used by Baker et al. (2004) came in explicit pairs (often connected as a single object), and the statistical correlations always held only between the items in these pairs. As explored in the General Discussion, this is a critical point for any study of statistical learning: Given the massive amount of incoming input in the real world, it is crucial to limit the populations over which statistical learning will apply (see also Fiser & Aslin, 2005).

[2] We thank Dick Aslin for providing 12 of these 24 shapes, as used in Fiser and Aslin (2001, 2002a).

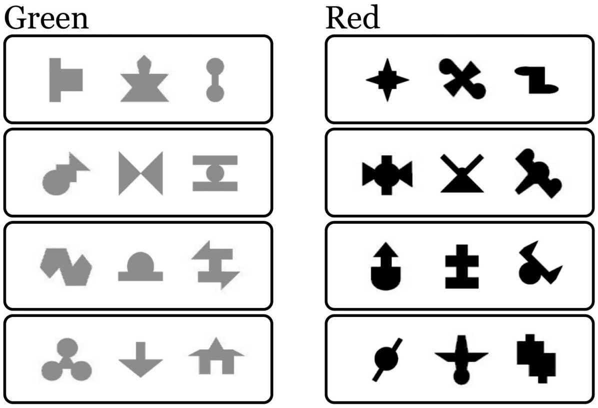

Figure 1. The 24 shapes used in the familiarization phase of all experiments, grouped into one possible set of triplets, divided by stream.

many other types of visual processing, including studies of implicit learning. We thus began this experiment without any a priori prediction about the results, but with the assumption that they would nevertheless serve to clarify the underlying nature and automaticity of VSL.

Method

Observers

Eight undergraduate students participated in exchange for course credit. In this and the four subsequent experiments, all observers reported normal or corrected-to-normal vision.

Apparatus

All displays were presented on a 17-in. (43.18-cm) NEC Multisync monitor attached to an Apple PowerMac G4 computer. Observers were positioned approximately 60 cm from the display, without head restraint, such that the display subtended approximately 29.4° by 21.9° of visual angle. (Stimulus sizes below are reported assuming this viewing angle.) The displays were presented with custom software written with the use of the VisionShell graphics libraries (Comtois, 2004).

Stimuli

Shapes. Over the course of the experiment, observers were presented with 24 novel shapes (see Figure 1), each subtending approximately 3.3°. Twelve of these shapes were identical to those used in the previous studies of Fiser and Aslin (2001, 2002a), and 12 were similar novel shapes constructed for this study. Full sets of the 24 shapes were generated in red, green, and black. During the experiment each shape appeared in isolation in the center of the display.

Familiarization streams. For each observer, 12 shapes were assigned to the red group, and the remaining 12 shapes were assigned to the green group. Within each color, the 12 shapes were further divided into four groups of three shapes (i.e., into four "triplets"). Separate temporal streams were first generated for each color, consisting of 24 repetitions of each triplet randomly intermixed. To manipulate attention (as described below), we also included in each stream 24 instances in which the third shape of a triplet was immediately repeated (e.g., *ABCCDEF*). Each stream thus included a total of 312 shapes. Within each stream, the randomized triplet order was constrained in two ways (using T_1 and T_2 as hypothetical triplets): No repeated triplets were allowed (e.g., . . .T_1T_1. . .), and no repeated pairs of triplets were allowed (e.g., . . .$T_1T_2T_1T_2$. . .). The implementation of these constraints was blind to target repetitions, such that a sequence of the form *ABCCABC* was still invalid. All triplet orders were derived ahead of time and used equally often for attended and unattended streams, counterbalanced across observers.

Interleaving. Though the red and green streams were constructed independently, they were randomly interleaved into one long stream during the stimulus presentation, such that each observer was familiarized with a single regular temporal sequence of 624 shapes (see Figure 2). This interleaving occurred by randomly sampling the two color streams in order and without replacement, with the single constraint that the remaining pool of shapes from one color could never exceed that of the other color by more than 6 shapes. Alternation between the two streams was counterbalanced such that one observer would attend to one of the streams, and another to the other stream. In addition, the sequence of alternation was reversed for two more observers, and attention was again counterbalanced. The resulting joint probabilities of various three-shape sequences can thus be computed both for each color stream independently, and for the final interleaved stream (see Table 1).

Procedure

Familiarization phase. Observers were seated in a dimly lit room. The experimenter first described the initial phase of the experiment via verbal instructions that observers could also follow in written form. Observers were told that they would see a sequence of red and green shapes, one at a time, in the center of a white background. Depending on the condition to which they had been randomly assigned, they were told to monitor one stream of shapes (either red or green, counterbalanced across observers) and to press a key whenever they observed an immediately subsequent repetition of a shape in that stream. Note that this task was not a simple one-back task in the final interleaved stream, because the target repetition could be interrupted by a variable number of shapes from the other color. This task served to bias attention to shapes of just one color. Observers began the initial phase of the experiment by pressing a key to begin the sequence. The shapes appeared at an SOA of 400 ms, with an interstimulus interval (ISI) of 200 ms. Responses in the repetition detection task were recorded to ensure that they completed the cover task, with responses made within 3,000 ms counted as correct repetition detections. Responses falling outside of this window were counted as false alarms.

Test phase. After the presentation of all 624 shapes (which took 4 min 10 s), the animation ended and the instructions for the test phase appeared.

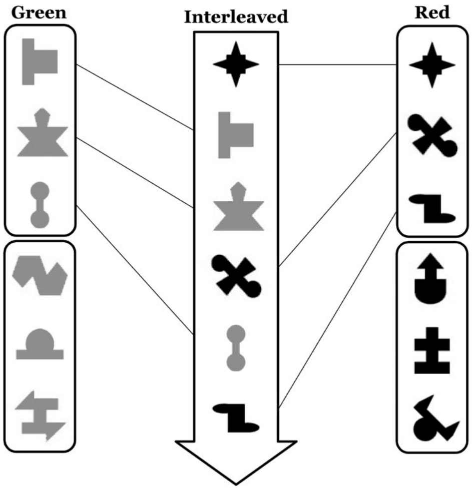

Figure 2. Example streams of triplets in both colors, randomly interleaved as in the actual displays.

Table 1
Joint Probabilities of Shapes in Each Stream in Isolation (as Used to Construct Stimuli) and Interleaved (as Shown to Observers)

			Without target		With target	
Stream	Isolation	Interleaved	Isolation	Interleaved	Isolation	Interleaved
p(any particular shape), e.g., *p*(A)	1/4 × 1/3	1/12 × 1/2	0.083	0.042	0.077	0.039
p(any third triplet shape), e.g., *p*(C)	1/4 × 1/3	1/12 × 1/2	0.083	0.042	0.096	0.048
p(any pair within a triplet), e.g., *p*(A, B)	1/12 × 1/1	1/24 × 1/2 × 1/1	0.083	0.021	0.077	0.020
p(any pair spanning triplets), e.g., *p*(C, G)	1/12 × 1/3	1/24 × 1/2 × 1/3	0.027	0.007	0.024	0.006
p(any given triplet), e.g., *p*(A, B, C)	1/12 × 1/1 × 1/1	1/24 × 1/2 × 1/1 × 1/2 × 1/1	0.083	0.010	0.077	0.010
p(any given nontriplet), e.g., *p*(B, C, G)	1/12 × 1/1 × 1/3	1/24 × 1/2 × 1/1 × 1/2 × 1/3	0.027	0.003	0.019	0.002
p(any foil sequence), e.g., *p*(A, E, I)	1/12 × 0/1 × 0/1	1/24 × 1/2 × 0/1 × 1/2 × 0/1	0	0	0	0

On each test trial observers viewed two 3-shape test sequences, each presented in the same manner (and with the same timing) as in the initial phase, temporally segmented by a 1,000-ms pause. All shapes during the test phase were drawn in black (instead of red or green, the colors in which these shapes were learned). One of the test sequences was a triplet from either the red or the green familiarization stream (e.g., *ABC*), although observers would have rarely encountered this triplet during familiarization without shapes from the other color interleaved into it. The second test sequence was a foil sequence constructed from three shapes of different triplets from that same color stream (e.g., *AEI*).

After the presentation of the two test-sequences on each trial, observers were instructed to simply press one of two keys in an unspeeded manner to indicate whether the first or second test-sequence seemed more familiar based on the initial animation, beyond familiarity with the individual shapes; no feedback was provided. In all cases, observers had seen the individual shapes in the triplet versus foil sequences an equal number of times, and so the sequences were only distinguished by the familiarity of their higher-order temporal structure. Each of the eight triplets from the initial phase (four of each color, *ABC, DEF, GHI, JKL*) was tested eight times, paired twice with each of four different foil sequences from that same color (*AEI, DHL, GKC, JBF*), for a total of 64 test trials (4 triplets × 2 colors × 4 foils × 2 repetitions), presented in a different random order for each observer. The order of triplets versus foils in individual test trials was randomized and counterbalanced for each observer. Observers' accuracy in discriminating triplets from foil sequences was used as the measure of statistical learning, as in Fiser and Aslin (2002a), and was calculated here separately for triplets initially encountered in the attended versus the unattended color.

Results and Discussion

All observers satisfactorily completed the cover task during the learning phase, detecting 71.4% of the repetitions in the attended color ($SD = 14.5\%$) and incorrectly responding to nontargets less than one time on average throughout the entire familiarization phase (range = 0–2 times). Accuracy in discriminating the triplets from the foil sequences in the 2IFC familiarity test phase is depicted in Figure 3A. Chance performance for these judgments would be 50% accuracy. Inspection of Figure 3A readily reveals the two primary results from this study. First, performance was better for attended than unattended shapes. Second, there was apparently no statistical learning for unattended shapes. These impressions were verified by submitting the data to three two-tailed planned comparisons. Learning of the statistical regularities among attended shapes was better than chance (59%), $t(7) = 2.99$, $p = .02$, $\eta_p^2 = .56$, but there was no learning for unattended shapes (49%), $t(7) = 1.00$, $p = .35$. Learning was significantly better for attended versus unattended shapes (a 10% difference), $t(7) = 3.42$, $p = .01$, $\eta_p^2 = .63$.

These results suggest that attention modulates and may even be necessary for VSL. However, because learning in the attended stream was reasonably close to chance performance and was much lower than is typically observed in similar paradigms (e.g., Fiser & Aslin, 2002a), it is possible that learning from the unattended stream would have been apparent with a more sensitive design. The overall low performance may have been due to several factors: (a) Our stimuli were presented at a faster rate than in all previous VSL experiments; (b) the two streams were presented in a temporally interleaved fashion, rather than in a unitary sequence as in all other studies; (c) success in our experiments required abstracting out the shape of the objects during the test phase because they always appeared in black; (d) the learning in our experiment had to occur completely incidentally, due to the novel cover task; and (e) learning had to occur in spite of the attentional demands imposed by the cover task. To assess the possibility of statistical learning in the unattended stream, however, we preferred not to alter any of these factors, because (as explored in the General Discussion) they each provide important theoretical insights into the nature of VSL.

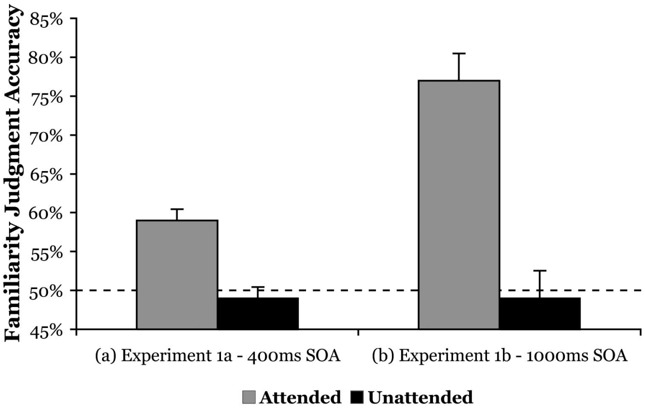

Figure 3. Accuracy in discriminating triplets from foil sequences in two-interval forced-choice familiarity judgments for (a) Experiment 1A and (b) Experiment 1B. Note the improvement in learning from 1A to 1B only in the attended stream. The dashed line indicates chance performance. Error bars correspond to within-subject standard errors. SOA = stimulus onset asynchrony.

Instead, in the next experiment we manipulated the reliability of statistical learning in a much more direct fashion.

Experiment 1B: Increased Exposure

This experiment was identical to Experiment 1A except that we quadrupled the exposure to each item by slowing the stimulus presentation rate. We expected that this would enhance any statistical learning. If statistical learning of both streams were improved in this manner, then this would suggest that the null result for unattended shapes in Experiment 1 was due to a floor effect. However, if learning of the attended shapes were enhanced in the absence of any learning for the unattended shapes, this would provide strong evidence that attention was required for VSL in this situation.

Method

Eight new undergraduate students participated in exchange for course credit. This experiment was identical to Experiment 1A, except that the SOA between shapes was extended to 1,000 ms to match previous studies, whereas the ISI was held constant at 200 ms. Thus each shape appeared for 800 ms instead of 200 ms as in Experiment 1A. The durations of the learning and test phases were accordingly extended, proportional to this increase in SOA (to 10 min 24 s and approximately 15 min, respectively).

Results and Discussion

All observers satisfactorily completed the cover task during the learning phase, on average detecting 88.5% of the repetitions in the attended color ($SD = 6.2\%$), and incorrectly responding to nontargets less than one time on average throughout the entire familiarization phase (range = 0–1 times). Accuracy in discriminating the triplets from the foil sequences in the 2IFC familiarity test phase is depicted in Figure 3B. Inspection of Figure 3B readily reveals that our speed manipulation enhanced statistical learning only for attended shapes (to 77%), whereas performance for unattended shapes remained at chance (49%). These impressions were verified by submitting the data to three two-tailed planned comparisons. Learning of the statistical regularities among attended shapes was better than chance, $t(7) = 5.85$, $p = .001$, $\eta_p^2 = .83$, but there was no learning for unattended shapes, $t(7) < 1$. Learning was again significantly better for attended versus unattended shapes (a 29% difference), $t(7) = 4.06$, $p = .005$, $\eta_p^2 = .70$.

Differences between Experiments 1A and 1B were assessed with a 2 (stream: attended vs. unattended) × 2 (experiment: 1A vs. 1B) mixed-design analysis of variance (ANOVA), with stream as a within-subjects factor and experiment as a between-subjects factor. There was a main effect of stream, with higher accuracy for triplets from the attended stream (68% vs. 49%), $F(1, 14) = 25.48$, $p < .001$, $\eta_p^2 = .65$, and also a main effect of experiment, with higher accuracy in Experiment 1B (63% vs. 54%), $F(1, 14) = 7.01$, $p = .019$, $\eta_p^2 = .33$. It is important to note that there was also a two-way interaction, reflecting the increase of accuracy in the attended stream without a commensurate boost (or, indeed, any boost at all) in the unattended stream, $F(1, 14) = 6.11$, $p = .027$, $\eta_p^2 = .30$. Experiments 1A and 1B thus provide strong evidence that attention is a critical mediator of VSL.

Experiment 2A: Maintaining Color Context

In the previous experiments, all of the shapes were displayed in red or green during familiarization but in black during test. This manipulation was included to avoid any continued attentional set effects for red versus green during the test phase, but, as explored below in the General Discussion, the maintenance of VSL despite this change has important implications for the degree of abstraction involved in VSL. This manipulation, however, may also have had an adverse effect on the expression of VSL. In Experiment 1B we failed to find any evidence for statistical learning of unattended regularities, despite the fact that the increased exposure in that experiment boosted learning of attended regularities. This could be due to the fact that any fragile statistical learning of the unattended regularities was depressed because of the removal of color context at test.

In this experiment we reinstated the color context at test in an effort to see if this manipulation would benefit VSL, especially for the unattended stream, perhaps by providing an additional retrieval cue. For example, shapes encountered in green during familiarization were also presented in green during the test phase (in both triplets and foils). We also retained the extended exposure from Experiment 1B, which, when combined with the maintained color context, provides an even more sensitive probe for whether there is any statistical learning for unattended regularities.

Method

Eight new undergraduate students participated in exchange for a monetary payment. This experiment was identical to Experiment 1B except that all shapes during the test phase were presented in the same color in which they were initially encountered during familiarization.

Results and Discussion

All observers satisfactorily completed the cover task during the learning phase, detecting 93.8% of the repetitions in the attended color ($SD = 12.8\%$), and incorrectly responding to nontargets less than one time on average throughout the entire familiarization phase (range: 0–1 time). Accuracy in discriminating triplets from foil sequences in the 2IFC familiarity test phase is depicted in Figure 4A. Inspection of this figure reveals that maintaining color context from familiarization to test had little effect. Statistical learning was again observed for attended shapes (69%), but this learning was no greater than in the previous experiment. Meanwhile, performance for unattended shapes remained exactly at chance (50%). These impressions were verified by submitting the data to three two-tailed planned comparisons. Learning of the statistical regularities among attended shapes was better than chance, $t(7) = 2.59$, $p = .036$, $\eta_p^2 = .49$, but there was no learning for unattended shapes, $t(7) < 1$. Learning was again significantly better for attended versus unattended shapes (a 19% difference), $t(7) = 2.60$, $p = .035$, $\eta_p^2 = .49$.

Experiment 2B: Swapping Color Context

The maintained color context in Experiment 2A did not noticeably improve statistical learning of either attended or unattended regularities. It remains possible, however, that the lack of any learning of unattended regularities in Experiment 2A is due to the

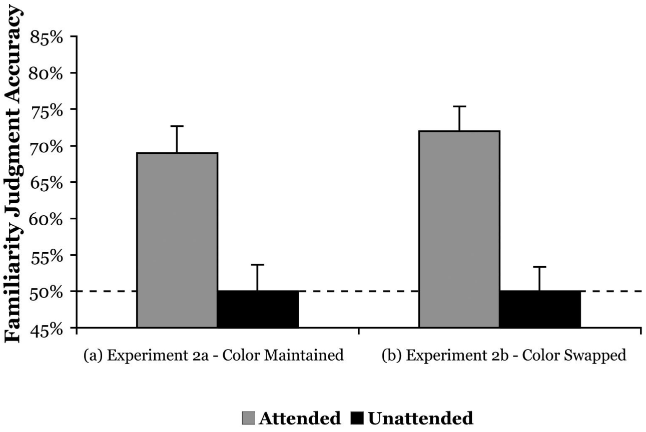

Figure 4. Accuracy in discriminating triplets from foil sequences in two-interval forced-choice familiarity judgments for (a) Experiment 2A and (b) Experiment 2B. Color did not affect statistical learning. The dashed line indicates chance performance. Error bars correspond to within-subject standard errors.

persistence of the attentional set from the familiarization phase. That is, attending only to green shapes during familiarization could, in principle, impair the recognition of red (unattended) regularities during test. This possibility, however, raises the unintuitive prediction that if the color contexts were swapped at test, with previously red shapes now drawn in green and vice versa, then the persistence of the attentional set from the familiarization phase might specifically benefit the expression of statistical learning for unattended regularities (and perhaps impair the expression of learning for attended regularities). This experiment thus constitutes another test of whether there is any statistical learning for unattended regularities at all. At the same time, this experiment, when combined with Experiments 1B and 2A, provides a test for whether the mental representations that result from statistical learning are sensitive to variation in extraneous surface features.

Method

Ten new undergraduate students participated in exchange for a monetary payment. This experiment was identical to Experiment 2A except that the color of the test shapes was reversed: shapes initially encountered in green during familiarization were presented in red during test and vice versa.

Results and Discussion

All observers satisfactorily completed the cover task during the learning phase, detecting 95.4% of the repetitions in the attended color (*SD* = 4.1%) and incorrectly responding to nontargets 1.1 times on average throughout the entire familiarization phase (range = 0–3 times). Accuracy in discriminating the triplets from the foil sequences in the 2IFC familiarity test phase is depicted in Figure 4B. Inspection of this figure reveals that swapping color context between familiarization and test had no appreciable effect. Statistical learning was observed for attended shapes (72%), whereas performance for unattended shapes again remained exactly at chance (50%). These impressions were verified by submitting the data to three two-tailed planned comparisons. Learning of the statistical regularities among attended shapes was better than chance, $t(9) = 3.50$, $p = .007$, $\eta_p^2 = .58$, but there was no learning for unattended shapes, $t(9) < 1$. Learning was again significantly better for attended versus unattended shapes (a 22% difference), $t(9) = 3.23$, $p = .010$, $\eta_p^2 = .54$.

Experiment 3: An Implicit Performance Measure

In the experiments reported thus far we followed many earlier studies of statistical learning in adults by using an explicit forced-choice familiarity judgment as the dependent measure. In a way, however, this is an odd sort of dependent measure to use, because it essentially asks observers to make an explicit judgment about implicitly learned relationships. (In fact, as emphasized in the General Discussion, our observers never reported awareness of the triplet-structure in the familiarization animation when asked.) Observers must thus respond on the basis of vague intuitive familiarity judgments, but such overt judgments may of course also be contaminated by other intuitions and response biases.

Moreover, these judgments may not truly reflect expression of statistical learning per se but rather a preference based on perceptual fluency (especially in the absence of explicit memory; Johnston, Hawley, & Elliott, 1991) as a result of the mere exposure effect (Bonnano & Stillings, 1986; Kunst-Wilson & Zajonc, 1980; Whittlesea & Price, 2001); conversely, it has been suggested that the mere exposure effect is itself the result of implicit memory (Seamon et al., 1995). In addition, some investigators seem to interpret any effects on overt familiarity judgments as indicating explicit rather than implicit learning (e.g., Baker et al., 2004). Finally, and most important, this dependent measure fails to address the question of what function VSL serves in the real world: Beyond resulting in vague implicit feelings of familiarity (perhaps only when probed), can statistical learning enhance other types of performance?

In this experiment, we attempted to test the automaticity of VSL by again manipulating attention to shapes of only one color during the familiarization phase, but then testing for learning using an implicit response-time (RT) measure. During each test trial of this experiment, observers viewed a brief, rapid, and unsegmented stream of previously encountered triplets, and we measured their response latency to detect a prespecified target shape. (All test shapes were again presented in black, as in Experiments 1A and 1B, because the previous experiments revealed no effect of color at test, and because retaining colors at test allows for the possibility of intrusive attentional set effects.) The target shape in each test trial could be the first, second, or third shape from a triplet, and we predicted that statistical learning would manifest itself in speeded responses to the later shapes in a triplet (the appearances of which may be primed by encountering the initial items in that triplet).

RT advantages of this type during statistical learning have been observed in three previous studies. In their study of learning, Baker et al. (2004) reported a time course function for the mapping of distractor–target pairs onto responses, and they observed that participants became faster at responding to high-frequency stimulus pairs versus low-frequency pairs. In a more complicated design, Hunt and Aslin (2001) used a modified serial reaction time task to investigate the time course of statistical learning. They constructed seven "words," each with three "syllables"; each syllable was a mapping of two of seven lights on a visual display to a given response. These seven words were repeated over many blocks, and as participants began to learn the words they became faster at responding to their second and third syllables. This statistical learning was apparently mediated only by learning the initial pair of syllables within each sequence, because there was an RT advantage for the second syllable over the first syllable, but no additional RT advantage for the third syllable over the second syllable (Hunt & Aslin, 2001), a pattern previously observed with familiarity judgments by Fiser and Aslin (2002a). Finally, in a study of temporal contextual cueing that served as the inspiration for using this new dependent measure, Olson and Chun (2001) demonstrated that repeated identical temporal contexts speeded target detection in a rapid temporal stream of letters and numbers.

Part of the reason we used an implicit RT measure of VSL in this experiment was because it may be more sensitive to statistical learning of unattended stimuli. Indeed, because unattended objects may in some circumstances fail to be consciously perceived at all (as in studies of inattentional blindness; e.g., Most et al., 2005), it is possible that observers did learn the statistical structure of the shapes in the unattended color stream, but that this learning was not available for conscious report via overt familiarity judgments. If this is the case, then we would predict statistical learning to be apparent for both attended and unattended shapes in this experiment. If selective attention is required for statistical learning, however, then we would expect RT evidence of learning for only the attended stream. Beyond questions of attention, however, we wanted to test for this type of VSL in a performance-based task in order to demonstrate that statistical learning is not an epiphenomenon, but can actually improve other aspects of visual processing.

Method

Observers

Twelve new undergraduate students participated in exchange for course credit.

Stimuli

The shapes, streams, and statistics were identical to those used in the previous experiments, except for one important control. In the previous experiments, the 24 within-color-stream repetitions that occurred during familiarization (and that served as targets for the cover task) were always repetitions of the third shape of the previous triplet. This was done to avoid interfering with the joint probabilities among shapes within each triplet, but it raises the possibility that observers were able to somehow extract the triplets because of the repetitions. In fact this seems unlikely for several reasons: (a) These repetition targets were relatively rare; (b) they appeared in both the attended and unattended streams; (c) the target repetitions would not fully demarcate a triplet, but could rather serve as a cue only to where one began or ended; and (d) as emphasized in the General Discussion, the overt cover task did not result in conscious detection of the triplet structure. Nevertheless, during the familiarization phase in this experiment, the 24 repetitions were sampled from either the first or last shape of triplets; in this way it was impossible for observers to determine the triplet structure only on the basis of the repetition locations. In addition, brute frequency information was identical for both the first and last shapes in both streams, and any processing benefit cannot be attributed to frequency effects.

Procedure

The learning phase was identical to that in Experiments 1B, 2A, and 2B (with a 1,000-ms SOA), and the test phase again involved all shapes drawn in black. However, the procedure used during the test phase was otherwise radically different from the previous experiments. At the beginning of each test trial, observers were presented with one of the 24 shapes from the learning phase (the shape was shown on an instructions screen that persisted until the observer pressed a key to continue). This shape served as the "test target" for that trial; observers were instructed to look for this shape during the ensuing test sequence, and to respond as quickly as possible with a keypress when it was detected. Each test trial stream began after a 1-s pause following the instructions screen and consisted of 24 shapes: two repetitions each of the four triplets from the same color-stream as the test target, in a random order. (Each test target thus appeared twice in the test sequence, but could never appear as either the first or last triplet.) These test sequences were presented more rapidly than the familiarization phase, with each shape presented for 200 ms followed by a 200-ms pause before the next shape. Target detection accuracy and RTs were recorded on each trial. Each of the 24 shapes in the familiarization phase served as a test target four times, for a total of 96 test trials presented in a random order.

Data Analysis

The test trials were divided into attended and unattended target groups, based on their original color in the familiarization phase. They were further divided according to the index of the test target in its triplet from the learning phase. This resulted in a 2 (color: attended vs. unattended) × 3 (intratriplet item position: first, second, or third) design, with a total of 32 RTs recorded for each of these six cells during the 96 trials. Response times were trimmed by removing responses falling 3 standard deviations outside each participant's individual mean (resulting in the removal of 1.3% of responses).

Results

As in the previous experiments, all observers satisfactorily completed the cover task during the learning phase, detecting 80.9% of the repetitions in the attended color (*SD* = 9.1%), and incorrectly responding to nontargets less than two times on average throughout the entire familiarization phase (range = 0–3 times). During the test phase, target detection accuracy was extremely good for both originally attended and unattended target shapes (96.6% and 96.2%, respectively). Test target detection accuracy did not differ based on stream (attended vs. unattended), $F(1, 11) < 1$, or intratriplet position (first vs. second vs. third), $F(2, 22) < 1$, and these factors did not interact, $F(2, 22) < 1$.

The RT data from the test phase are presented in Figures 5A (for attended shapes) and 5B (for unattended shapes). As is clear from these graphs, the RTs revealed statistical learning only for attended shapes. This impression was confirmed in several analyses, beginning with a 2 (stream: attended vs. unattended) × 3 (triplet position: first, second, or third) repeated measures ANOVA. There was no main effect of stream, $F(1, 11) < 1$, or of triplet position, $F(2, 22) = 2.25$, $p = .13$. It is important to note, however, that there was a reliable interaction between stream and position, $F(2, 22) = 3.49$, $p = .048$, $\eta_p^2 = .24$, suggesting that the hypothesized decrease in RTs for later positions in learned triplets was specific to targets from the attended stream. Planned follow-up comparisons confirmed this interpretation by revealing a main effect of position for targets from the attended stream, $F(2, 22) = 5.71$, $p = .01$, $\eta_p^2 = .34$, but not for targets from the unattended stream, $F(2, 22) < 1$. In addition, and in contrast to previous work (Hunt & Aslin, 2001; Fiser & Aslin, 2002a), there was evidence for learning within attended triplets between Positions 1 and 2, $t(11) = 1.94$,

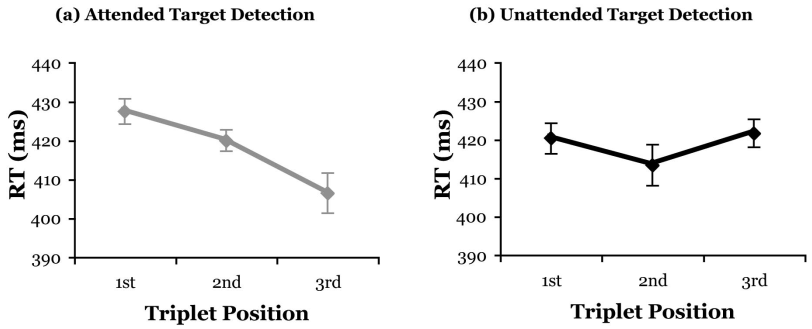

Figure 5. Response times to detect a prespecified target in the rapid test streams of Experiment 3 for (a) attended shapes and (b) unattended shapes. This implicit task revealed robust statistical learning only for attended shapes. *x*-axis: target position in its original triplet during familiarization. Error bars correspond to within-subject standard errors. RT = response time.

$p = .039$ (one-tailed), $\eta_p^2 = .25$; and Positions 2 and 3, $t(11) = 1.97$, $p = .038$ (one-tailed), $\eta_p^2 = .26$.

Discussion

These results confirm that VSL can have additional effects on (and can be assessed by measuring) other aspects of visual processing, without an overt familiarity judgment. Moreover, these results converge on the conclusion that attention is required for VSL, and they demonstrate that the failure to observe learning in the unattended stream in earlier experiments was not simply due to the explicit familiarity task.

Because the test displays themselves had statistical structure, and indeed were quite similar to the familiarization displays of Fiser and Aslin (2002a), one might expect that statistical learning (now of both sets of triplets, because there was no attention manipulation or color difference) would occur during the test phase in this experiment (cf. Onishi, Chambers, & Fisher, 2002). To assess this possibility, the RTs were divided into two blocks corresponding to the first and second halves of test. If learning occurred during test, one might expect an overall effect of block on RT or perhaps an interaction of block with position, such that shapes in the third intratriplet position get faster and faster, whereas the initial shapes do not see an equal boost. Similarly, an interaction between block and stream would suggest a selective learning advantage for attended or unattended triplets. However, a 2 (stream) × 3 (position) × 2 (block) repeated measures ANOVA revealed no main effect of block, $F(1, 11) < 1$, and no interactions of block with stream or position (ps $> .23$). This suggests that in fact no learning occurred during the test phase. This may be due to the fact that there are robust primacy effects in VSL: Statistical learning (or the lack thereof) seems to be determined primarily by the statistical information encountered in the initial phases of familiarization, such that statistical information followed by structureless noise will be learned, but noise followed by structure will not (Catena, Scholl, & Isola, 2005). Things are more complicated in the present study because some shapes were unattended during familiarization but later (presumably) attended at test, but it nevertheless appears that observers did not extract the statistical information inherent in these shapes during test.

General Discussion

The five experiments reported here collectively provide strong evidence that VSL is constrained by selective attention. We first evaluate the implications of these results for the automaticity of VSL and relate these results to similar research on other types of implicit learning. We then discuss several other ways in which these results enhance our understanding of the underlying nature of statistical learning, and we outline a few important opportunities for additional research into these questions. We conclude by suggesting that VSL both is and is not automatic in important ways.

Attention

Each of the five experiments reported in this article provides evidence that VSL of temporal sequences is constrained by the allocation of attention. In Experiment 1A, attention was manipulated by giving observers a cover task involving just one of the two colors in which the stimuli were presented, and learning of the statistical regularities was later assessed via familiarity judgments between pairs of three-item sequences. The shapes used in all test sequences had been observed an equal number of times, but their higher-level sequential structure differed: Each trial pitted a triplet from the familiarization phase against a foil sequence of shapes from the same color-stream that had not previously occurred in that order. After just 6 min of familiarization, observers reliably discriminated the triplets from foils, but only for shapes presented in the attended color.

Additional exposure to the familiarization streams in Experiment 1B succeeded in boosting learning for the attended stream but had no effect on unattended items. The colors of the items at test in Experiments 1A and 1B were always black, leaving open the possibility that this change had obscured learning of the unattended regularities. Experiments 2A and 2B ruled out this possibility, however, by demonstrating that the color of the items at test did not matter: there were no differences in the degree of learning (in the attended stream) or the lack thereof (in the unattended stream) when the test items were drawn in black (in Experiment 1B), in their original colors (in Experiment 2A), or when the colors were swapped at test (in Experiment 2B).

Experiment 3 then used a novel implicit performance measure based on RT, and confirmed that the lack of learning for unattended items was not simply due to the fact that we had been using an overt familiarity judgment. Overall we failed to find any hint of statistical learning from unattended shapes in any of the five experiments, despite the robust statistical learning of attended shapes in some of the most demanding conditions yet tested (as discussed below).

This reliable effect of selective attention in VSL is consistent with demonstrations of a role for selective attention in other types of implicit learning. Though dual tasks in general do not necessarily interfere with implicit learning of other types of sequences (e.g., Cohen et al., 1990; Frensch et al., 1998; Stadler, 1995), tasks that require selective attention to only certain stimuli in a sequence do attenuate or eliminate implicit learning in other contexts involving perceptual or motor correlations (Jiménez & Méndez, 1999; Nissen & Bullemer, 1987; though cf. Willingham, 1999, who suggests that some such tasks may reflect primarily motor-based learning). The present results extend this effect of selective attention to a situation involving VSL (Fiser & Aslin, 2002a) with no motor component, involving learning that occurred over the span of only a few minutes (vs. the more than 30,000 trials used by Jiménez & Méndez, 1999), and employing the same types of probabilistic sequences used in most recent temporal statistical learning studies (e.g., Saffran et al., 1996; cf. the entirely deterministic sequence used in Nissen & Bullemer, 1987).

It is also possible that the attentional effects observed in our experiments are related to other types of selectivity in statistical learning. For example, participants in a recent experiment by Fiser et al. (2004) repeatedly viewed one shape that moved toward and then passed behind an occluder, followed by two different shapes that emerged from behind the occluder. Because of the dynamics of the shapes' motions, only one of the two emerging shapes appeared to be perceptually linked to the initially occluded shape, appearing as a continued state of the same persisting object. In this situation, even though both of the emerging shapes were equally statistically predicted by the initial occluding shape, participants had a bias to learn the statistics of only those shape pairs that were perceptually bound into the same enduring objects on the basis of their motion patterns. Although this study did not involve any direct measurement of attention, it is possible that participants attended more to the disoccluding shape that was seen as the same object that was initially occluded—and that this attentional asymmetry could have contributed to the resulting difference in VSL. This interpretation is supported by the fact that there was no bias in subjects' overt eye movements while viewing the displays in this other study.

It remains possible, of course, that other situations may still allow for statistical learning of unattended items, in nonselective situations. A critical aspect of these experiments, and one that is highly relevant to real-world perception, is that observers always had to select some stimuli (in the color cued by the cover task) against others (in the other color). Thus the selection may have involved (a) devoting extra attentional resources to shapes of the cued color, and/or (b) actively ignoring shapes of the uncued color, which then gated "normal" statistical learning of the remaining objects. This type of selection via active inhibitory ignoring has been observed in several other attentional phenomena, such as inattentional blindness (Most et al., 2001) and multiple-object tracking (Ogawa & Yagi, 2003; Pylyshyn, in press). In this way, it remains possible that VSL could operate over unattended shape sequences if they were not being actively ignored, but of course this possibility is difficult to assess because without an overt task, observers are likely to attend to the shapes.[3]

Another possibility is that VSL for unattended objects could be revealed with a more sensitive measure, but the present studies already rule out one such possibility that has played a role in related types of implicit learning. In the domain of contextual cueing, for example, Jiang and Leung (2005) recently developed an especially sensitive measure for determining whether ignored visual contexts were still implicitly learned over time. They began by replicating the results of Jiang and Chun (2001), showing that predictive contexts do not facilitate visual search when those contexts are learned in an unattended color. After learning, however, they then unexpectedly swapped some of the unattended predictive contexts into the attended color, and some of the attended predictive contexts into the unattended color. Surprisingly, previously unattended predictive contexts that had not facilitated search now began to facilitate search immediately. Likewise, previously facilitative contexts, when switched to the unattended color, no longer facilitated search. Note, however, that Experiment 2B of the present study effectively rules out this possibility in the context of VSL. When we swapped the colors at test, we obtained the opposite result from Jiang and Leung (2005): Learning of regularities involving previously attended shapes continued to be expressed when swapped to the previously unattended color, whereas no learning of regularities involving previously unattended shapes was expressed even when swapped to the previously attended color.

In the remainder of this article we first stress three other aspects of these results that help to clarify the underlying nature of statistical learning (and to demonstrate its flexibility), and we conclude by drawing several implications for the automaticity of this type of visual processing.

Learning Through Noise?

Our experiments used interleaved familiarization streams simply because this was an effective way to manipulate attention, but this interleaved structure is also independently interesting and potentially important. In particular, it suggests that VSL is possible even when the related items (i.e., the three items in each triplet) do not occur in immediate succession, but are rather interrupted by other irrelevant items (i.e., the intervening items from the other color-stream). Nonadjacent dependencies of this type have featured prominently in recent research in the auditory domain because many salient aspects of linguistic structure involve long-distance dependencies. Accordingly, recent research with both adults and nonhuman primates has indicated that auditory statisti-

---

[3] One way to investigate this would be to bias learning toward only some features of the familiarization streams (e.g., shape) without selecting against any other shapes, and to later test for statistical learning of unattended features such as color that were also independently statistically structured. This would involve attention to objects, but not to the tested featural dimensions, and we are currently pursuing this possibility in "multidimensional" VSL experiments in which several featural dimensions vary independently.

cal learning is not restricted to temporally contiguous relationships, but can also pick out nonadjacent relations in sequences of pseudospeech sounds and musical tones (Creel et al., 2004; Newport & Aslin, 2004; Newport et al., 2004). However, there do appear to be constraints on this ability. For example, only statistical structure between nonadjacent phonetic segments (consonants or vowels) appears to be learnable for humans (although see Bonatti et al., 2005, for evidence that vowels are poorly learned), and this ability may be restricted to "every-other-item" patterns of nonadjacency, such that there is always a single intervening item between the statistically related sounds.

Though the motivation for these auditory statistical learning experiments derives largely from linguistic considerations, the ability to learn through intervening noise seems equally important for visual perception. People are constantly bombarded with noise in space and time that needs to be segregated in order to extract a coherent representation of the world, and people rarely encounter a sequence of relevant stimuli without any interruptions. Consider driving, for example, in which drivers must constantly keep track of the road ahead, despite frequent glances to the rearview mirror, the radio, or other passengers. Thus, if VSL could not operate through noise, it would be of little utility in the real world. Our present results are suggestive that such *interrupted statistical learning* is possible, because all of our familiarization displays involved interleaved displays. Additional research directed at this question is clearly required, however, because the interleaving in our displays was random, and thus some triplets did occasionally occur without any interruptions just by chance. At the same time, however, the successful VSL through random interleaving suggests that every-other-one constraints may not apply in the visual domain (in which indeed they would make little sense). For additional experiments that investigate this issue directly, along with discussion of similar effects in other types of implicit learning (e.g., contextual cueing and artificial grammar learning), see Jungé, Turk-Browne, and Scholl (2005).

Abstraction in Statistical Learning

Another important aspect of our results is the fact that the shapes were rendered in black during the test phase of Experiments 1A, 1B, and 3, despite the fact that they had originally been encountered in either red or green. This was another design detail that was implemented in the service of the attentional manipulation, because it eliminated at test what might otherwise have been an overt cue (i.e., color) to whether the shapes were encountered in the attended or unattended streams. As with the interleaved nature of the streams, however, this manipulation is also independently important. To our knowledge, all previous studies of VSL have used shapes during the test phase that were identical to the shapes used during familiarization. Thus it is possible in these experiments that statistical learning only transfers to stimuli that are identical to earlier experiences. Of course, this limitation would also dramatically reduce the relevance of such processing to real world perception, in which stimuli are always changing because of viewpoint or perspective shifts, lighting changes, and so forth. In other words, some type of abstraction from local stimulus details must exist for VSL to be useful.

Our experiments suggest that at least some such abstraction is possible, because statistical learning of the shapes in the attended color occurred despite the change from red (or green) to black shapes at test (in Experiments 1A, 1B, and 3), and indeed, despite swapping the colors entirely (in Experiment 2B). These results collectively suggest that VSL constructs abstracted representations that are then invariant to changes in extraneous surface features (i.e., color, in these experiments). This is a theoretically important result and is the first demonstration that the mental representations extracted during VSL encode only some stimulus properties. These results are consistent with other recent evidence from auditory statistical learning of syllable sequences, in which learning can be expressed despite changes in prosody (Thiessen, Hill, & Saffran, 2005; Thiessen & Saffran, 2003). These findings thus raise a new question about which features are and are not encoded during VSL (in other words, about what is learned during statistical learning). Previous studies were consistent with the possibility that retinally accurate memory traces were extracted during VSL, but the current results suggest that this processing is more nuanced, encoding only some stimulus properties and discarding others. Additional experiments are underway to explore the generality and constraints on such abstraction and transfer.

The Speed of Statistical Processing

One final note of methodological import involves the "speed limit" of VSL. The shapes in Experiment 1A were presented 2.5 times faster than in previous studies of temporal VSL (e.g., Fiser & Aslin, 2002a), but their statistical structure was still extracted in the attended color. This suggests that VSL is efficient enough to operate in real-time contexts and can quickly adapt to the often rapid pace of natural visual events. It would be interesting for future research to explore these temporal constraints directly: How fast can stimulus sequences occur in order for their statistical structure to be extracted? At what temporal separation are two events considered to be the same event? On the other end, what is the longest gap over which stimuli can be strung together and still have their statistical relationship be automatically computed (and how does this interact with selective attention)?

Conclusions: The Automaticity of Statistical Learning

The results of our attention manipulation and several other aspects of our study suggest that VSL both is and is not automatic, in different senses. On the one hand, the discovery that selective attention is required for VSL in this context clearly argues against a strong form of automaticity in which the visual system picks up the statistical structure in a reflexive and preattentive manner, as a product of mere exposure, and regardless of what stimuli observers may be selecting (or selecting against) via endogenous attention. (It is worth remembering that this was an entirely possible pattern of results, given that previous discussions have described statistical learning as a low-level implicit process.) Instead, it seems that VSL occurs for only some information in our local environment, and that attention is at least one of the ways that this choice is made.

In fact, the necessity of attention for VSL seems less surprising when we consider the character of our natural environment, and the fact that any statistical process must operate over some specified population. As noted in the opening paragraph of this article, one challenge faced by visual processing is that we are perceptually confronted with far too much sensory input to fully process, and thus it seems computationally unlikely that VSL would operate over every possible stimulus we encounter (see also Fiser & Aslin, 2005). Rather, VSL may operate only over populations of objects and events that are important to us, and in this sense our results suggest that the populations over which statistical computations will occur are not chosen in an automatic data-driven fashion, but rather can be influenced by the application of selective attention. Such selection is important not only for determining the contents of our conscious experience, but also for determining the stimuli over which other lower-level processes will operate.

At the same time, other aspects of our results continue to support the view that VSL is automatic in several other senses. First, recall that previous statistical learning studies involved mere exposure to the familiarization stimuli in which observers simply watched (or listened to) the stimuli with no competing demands. As part of this (odd) experience of simply watching, however, it remains possible that observers at least implicitly tried to understand what they were seeing. In contrast, observers in all of our experiments were always engaged by a cover task during familiarization that required them to process the sequences in a different way (to detect repetitions rather than to extract statistical information per se). It is notable and important that our observers learned the statistical patterns in the familiarization phase for attended objects despite this competing demand, even though they were actively trying to process the stimuli in another way. A second and even more direct demonstration of (a type of) automaticity in VSL is the fact that no observers indicated during careful debriefing in Experiment 3 that they were aware of the structure in the displays. In other words, as is the case with all forms of implicit learning (for a review, see Stadler & Frensch, 1997), observers had learned something that they did not know they had learned.

These factors clarify the ways in which VSL both is and is not automatic: The determination of the populations over which VSL will operate does not appear to be automatic (rather it is influenced by and perhaps even requires attention), but the actual statistical operations themselves appear to proceed despite the fact that the relevant stimuli are presented extremely quickly while interleaved in noise, despite the fact that observers have no intent to extract this structure (and indeed are engaged in a competing overt task), and despite the fact that this processing does not result in any direct awareness of the structure that is being learned.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324.

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

Bates, E., & Elman, J. (1996, December 13). Learning rediscovered. *Science, 274,* 1849–1850.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16,* 451–459.

Bonnano, G. A., & Stillings, N. A. (1986). Preference, familiarity and recognition after repeated brief exposures to random geometric shapes. *American Journal of Psychology, 99,* 403–415.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7,* 308–313.

Catena, J. A., Scholl, B. J., & Isola, P. J. (2005). *The onset and offset of visual statistical learning.* Manuscript submitted for publication.

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

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 224–234.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Comtois, R. (2004). Vision Shell PPC [Computer graphics library]. Cambridge, MA: Author.

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,* 1119–1130.

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

Fiser, J. Z., & 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. Z., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA, 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. Z., Scholl, B. J., & Aslin, R. N. (2004). Perception of object trajectories during occlusion constrains statistical learning of visual features [Abstract]. *Journal of Vision, 4*(8), 189.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of secondary tone-counting on implicit learning in the serial reaction time task. *Psychological Research, 61,* 83–98.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53–B64.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences, 1,* 115–121.

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., & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54,* 1105–1124.

Jiang, Y., & Leung, A. W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin & Review, 12,* 100–106.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 236–259.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37,* 43–75.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 210–223.

Jungé, J. A., Turk-Browne, N. B., & Scholl, B. J. (2005). Visual statistical learning through intervening noise [Abstract]. *Journal of Vision, 5*(8), 421.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35–B42.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980, February 1). Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557–558.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Cambridge, MA: MIT Press.

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review, 112,* 217–242.

Most, S. B., Simons, D. J., Scholl, B. J., Jiminez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science, 12*(1), 9–17.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.

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.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32.

Ogawa, H., & Yagi, A. (2003). Priming effects in multiple object tracking: An implicit encoding based on global spatiotemporal information [Abstract]. *Journal of Vision, 3*(9), 339.

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

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*(1), B13–B23.

Pashler, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002, October 18). Signal-driven computations in speech processing. *Science, 298,* 604–607.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133,* 573–583.

Pylyshyn, Z. W. (in press). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition.*

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–105.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39,* 484–494.

Seamon, J. G., Williams, P. C., Crowley, M. J., Kim, I. J., Langer, S. A., Orne, P. J., & Wishengrad, D. L. (1995). The mere exposure effect is based on implicit memory: Effects of stimulus type, encoding conditions, and number of exposures on recognition and affect judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 711–721.

Seidenberg, M. S. (1997, March 14). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599–1604.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 674–685.

Stadler, M. A., & Frensch, P. A. (Eds.). (1997). *Handbook of implicit learning.* London: Sage.

Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7,* 53–71.

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.

Whittlesea, B. W. A., & Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition, 29,* 234–246.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition, 27,* 561–572.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences, 8,* 451–456.

Received March 22, 2005
Revision received June 28, 2005
Accepted June 30, 2005 ■