Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness

Nicholas B. Turk-Browne, Brian J. Scholl, Marvin M. Chun, and Marcia K. Johnson

Abstract

■ Our environment contains regularities distributed in space and time that can be detected by way of statistical learning. This unsupervised learning occurs without intent or awareness, but little is known about how it relates to other types of learning, how it affects perceptual processing, and how quickly it can occur. Here we use fMRI during statistical learning to explore these questions. Participants viewed statistically structured versus unstructured sequences of shapes while performing a task unrelated to the structure. Robust neural responses to statistical structure were observed, and these responses were notable in four ways: First, responses to structure were observed in the striatum and medial temporal lobe, suggesting that statistical learning may be related to other forms of associative learning and relational memory. Second, statistical regularities yielded greater activation in category-specific visual regions (object-selective lateral occipital cortex and word-selective ventral occipito-temporal cortex), demonstrating that these regions are sensitive to information distributed in time. Third, evidence of learning emerged early during familiarization, showing that statistical learning can operate very quickly and with little exposure. Finally, neural signatures of learning were dissociable from subsequent explicit familiarity, suggesting that learning can occur in the absence of awareness. Overall, our findings help elucidate the underlying nature of statistical learning. ■

INTRODUCTION

Our sensory environments are full of regularities distributed in space and time. For example, the syllable /sci/ is more likely to be followed by /ence/ than by /on/ in English speech; a microwave is more likely to be found near a stove than a furnace; and passing through a metal detector precedes getting on an airplane but not entering a shower. By being sensitive to this structure, we can acquire higher-order primitives—words, scene schemas, and event scripts.

However, such relationships are embedded within complex and continuous environments. Statistical learning is a way of acquiring structure in such situations, resulting in segmented ''units.'' In the auditory domain, for example, preverbal infants who are exposed to a continuous pseudospeech stream for 2 min can learn that some syllables are more likely to co-occur than others, providing a bootstrapping mechanism by which ''words'' can be identified (e.g., Saffran, Aslin, & Newport, 1996). In the visual domain, adult observers are sensitive to contingencies between shapes in temporal sequences (e.g., Turk-Browne, Jungé, & Scholl, 2005; Fiser & Aslin, 2002) and spatial configurations (e.g., Fiser & Aslin, 2001; Chun & Jiang, 1999). Such statistical learning may help define visual ''objects'' in time and/or space (Turk-Browne & Scholl, in press) and at multiple levels of complexity: binding features within objects (Turk-Browne, Isola, Scholl, & Treat, 2008), grouping objects into hierarchies (e.g., Orbán, Fiser, Aslin, & Lengyel, 2008; Fiser & Aslin, 2005), and abstracting exemplars to categories (Brady & Oliva, 2008).

Studies of statistical learning have explored its power and flexibility, demonstrating for example that it can occur in multiple sensory modalities (Conway & Christiansen, 2005, 2006), in complex dynamic displays (Fiser, Scholl, & Aslin, 2007), over nonadjacent syllables (Pacton & Perruchet, 2008; Newport & Aslin, 2004), despite cover tasks (Turk-Browne et al., 2005), and without conscious awareness of the regularities (Turk-Browne et al., 2005; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Less is understood, however, about the underlying perceptual and cognitive processes that contribute to statistical learning. Indeed, statistical learning is often discussed as a distinct category of learning (see Perruchet & Pacton, 2006) but may in fact have roots in other forms of learning and memory. Here we assessed these processes incidentally by measuring visual statistical learning with fMRI, without requiring the measurement of explicit knowledge in a separate behavioral post-test, which is the typical measure of statistical learning.

To our knowledge, there has been no investigation focused on the neural foundations of visual statistical learning of this sort. In the auditory domain, one study

Yale University

has examined the neural basis of how statistical and prosodic cues are integrated during word segmentation (McNealy, Mazziotta, & Dapretto, 2006), and another study has examined the time course of learning with ERPs and their relationship to familiarity (Abla, Katahira, & Okanoya, 2008). There have also been studies of the neural basis of other forms of learning, including artificial grammar learning (AGL; e.g., Skosnik et al., 2002), classification learning (e.g., Poldrack et al., 2001), and motor sequence learning (e.g., Toni, Krams, Turner, & Passingham, 1998; Grafton, Hazeltine, & Ivry, 1995). Thus, one goal of the present study is to empirically examine how these other forms of learning relate to statistical learning (for a theoretical discussion, see Perruchet & Pacton, 2006). Note, however, that statistical learning can be distinguished from these other forms of learning in at least two ways: (1) the output of statistical learning consists of stimulus-specific associations rather than abstract rules, probabilistic category labels, or motor programs; and (2) these associations create discrete units out of an otherwise undifferentiable input stream (i.e., regularities are demarcated only by statistics)—as opposed to AGL in which words are explicitly segmented, or classification learning in which associations are formed over cues and outcomes in discrete trials.

Much of the interest in statistical learning derives from its implicit nature: It can occur without intent or even knowledge that there are underlying regularities (Turk-Browne et al., 2005; Saffran et al., 1997). Indeed, a mechanism that relied on deliberate inference would be ill suited for learning at early stages of development—or perhaps at any stage, given the great number of potential regularities in the environment to assimilate. Nevertheless, studies of statistical learning have sometimes found it difficult to isolate implicit contributions, especially because most (adult) studies involve passive viewing during learning and explicit familiarity judgments at test (e.g., Turk-Browne et al., 2008; Fiser & Aslin, 2002). Such measures are conventional largely for historical reasons: This incarnation of statistical learning evolved from studies of infant cognition, wherein testing often occurs after a period of familiarization (e.g., Saffran et al., 1996) or habituation (e.g., Kirkham, Slemmer, & Johnson, 2002). Even many behavioral ''implicit'' measures (e.g., Turk-Browne & Scholl, in press; Turk-Browne et al., 2005) require a separate task at test (cf. Baker, Olson, & Behrmann, 2004). With neuroimaging, in contrast, we can explore the learning process itself, while observers are engaged in an orthogonal task.

We presented observers with short blocks of novel shapes appearing one at a time in a continuous stream. *Structured* blocks contained deterministic subsequences of shapes that only existed in terms of the higher transitional probabilities between shapes within a subsequence than between shapes spanning two different subsequences. *Random* blocks lacked this structure but were otherwise identical. Thus, any difference between these block types observed with fMRI must reflect sensitivity to the statistical structure. We explored four questions: (1) How does statistical learning relate to other forms of learning and memory? (2) What are the consequences of statistical learning for visual processing? (3) How fast and efficient is statistical learning? (4) What is the relationship between incidental statistical learning and subsequent explicit familiarity? In the context of our study, we operationalized these questions by examining, respectively, the extent to which statistical learning is mediated by the same brain systems as other forms of learning and memory, whether statistical learning modulates processing in category-specific ventral visual regions, how quickly sensitivity to structure can be observed, and whether neural evidence of learning can be dissociated from familiarity judgments.

METHODS

Participants

Sixteen naive observers (nine females; mean age = 23 years) participated in one fMRI session for monetary compensation. All were right-handed with normal or corrected-to-normal vision.

Stimuli

The stimuli consisted of 12 glyphs from the Sabaean alphabet (an ancient Semitic language) and 12 glyphs from the Ndjuka syllabary (a creole from Suriname). Images of glyphs were generated from fonts downloaded at www.omniglot.com. For each observer, the 24 glyphs were randomly assigned without replacement to either the structured set or the random set. The use of two alphabets helped increase shape discriminability, and the fact that structured and random sets randomly contained items from both alphabets helped prevent observers from treating the sets as categorically distinct. Each glyph subtended roughly 6.8°, appearing in black on a medium gray background (Figure 1A). A small blue dot was superimposed on the screen throughout each block to help observers stay fixated.

Similar to previous studies of statistical learning (e.g., Fiser & Aslin, 2002; Saffran et al., 1996), the structured blocks were constructed by assigning without replacement each of the 12 glyphs in the structured set to one of four ''triplets'': a subsequence of three glyphs that always appeared in the same order (Figure 1B). Each block consisted of one presentation of each triplet (Figure 1D). The order of triplets in each block was randomized but was not repeated in later blocks.

The use of deterministic triplets adds positional structure to the blocks: For example, the first glyph in a triplet only appeared in positions 1/4/7/10. Indeed, sequence learning can be supported by both item-to-item and item-to-position associations (e.g., Young, 1968). To control for this, the (pseudo)random blocks were

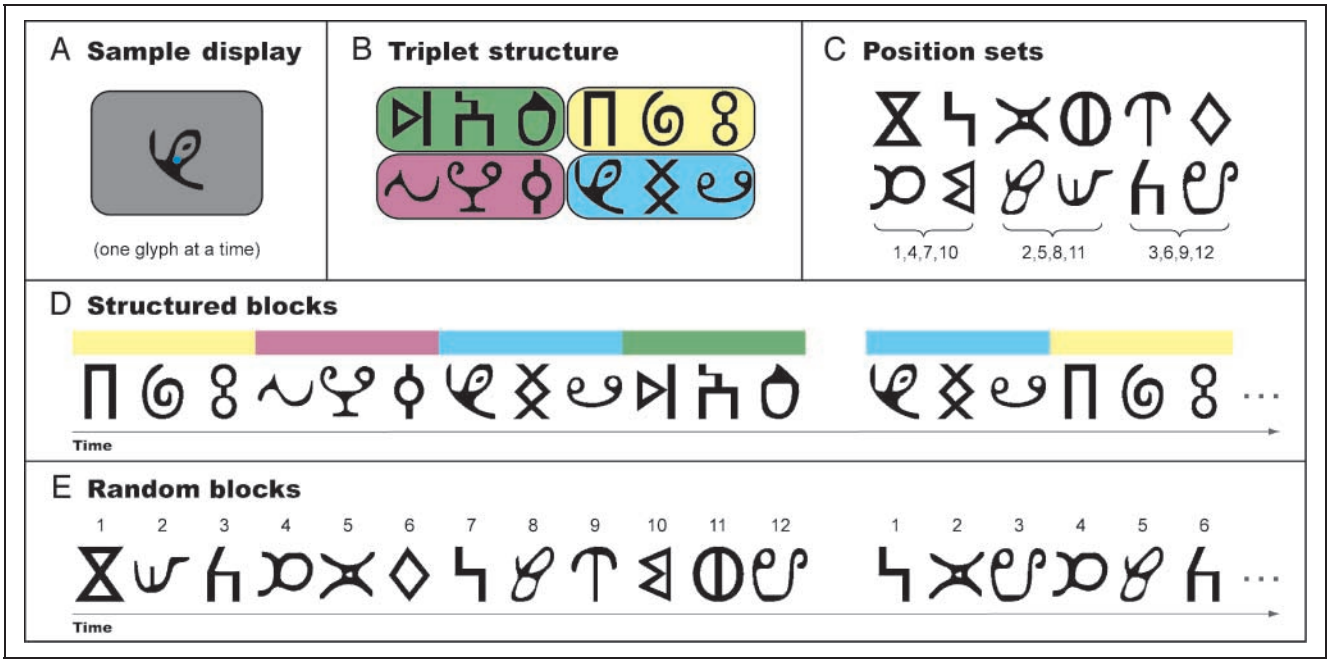

Figure 1. Stimuli and design. (A) Sample display of one glyph. (B) Twelve randomly assigned glyphs in four ''triplets.'' (C) The remaining 12 glyphs in three ''position sets.'' (D) Example structured block containing one presentation of each triplet in random order. (E) Example random block with glyphs constrained to certain positions.

constructed by assigning without replacement the 12 glyphs in the random set to one of four position sets (Figure 1C). That is, each glyph appeared once per block and only ever in positions 1/4/7/10, 2/5/8/11, or 3/6/9/12 (Figure 1E). In addition, because every glyph appeared once in each structured or random block, items near the end of blocks were more predictable if observers learned the stimulus sets; however, this was equally true for both block types. Thus, other than the lack of triplets, the random blocks were identical to the structured blocks in all respects (including overall novelty of individual glyphs and positional structure)—and because the assignment of glyphs to sets was randomized, any systematic neural differences between block types must therefore reflect sensitivity to the differential transitional probabilities.

Foil stimuli in a behavioral familiarity post-test were constructed from the structured set to mimic prior studies of statistical learning (e.g., Turk-Browne et al., 2005; Fiser & Aslin, 2002): Each of the 12 glyphs was assigned without replacement to one of four new ''foil'' subsequences. Every foil was constructed of one glyph from each position set (e.g., 1/5/9) and thus could only be distinguished from the triplets because the transitional probabilities between foil glyphs were zero based on the familiarization.

Procedure

As their task, observers used a button box to respond to ''jiggles'' (rapid motion of the current glyph to the left and right of fixation for 200 msec). Jiggle targets occurred once or twice per block at random intervals and an equal number of times in the two block types. There was a short practice run with unstructured blocks of line drawings during which observers practiced the task.

After anatomical scans, observers completed one run of the jiggle task containing structured and random blocks of glyphs. Due to technical difficulties, button responses from two participants were not recorded, but visual inspection revealed that they were performing the task. During analysis, jiggle detection response times more than three standard deviations greater than the mean were excluded as outliers (resulting in the removal of 1.8% of responses). There were twelve 16-sec blocks of each type, presented in an alternating manner for a total of 24 blocks. Typical statistical learning experiments use a continuous familiarization phase without blocks and do not contain a random control condition; rather, learning is assessed off-line after familiarization by testing regularities from familiarization against recombinations of the same elements. Our study, however, used an alternating structured/random block design due to practical challenges in interpreting fMRI data; in particular, comparisons between conditions that are spread across runs (or even that occur at different points within one run) can be easily confounded by head motion, signal drift, and/or changes in arousal. Although the block structure provided a form of segmentation that could facilitate learning, any learning related to the boundaries between blocks per se would have

existed in the random blocks as well due to the positional constraints.

Block order (i.e., whether the run began with a structured or random block) was counterbalanced across observers. In each block, the fixation dot appeared 1 sec prior to the first glyph, and then each of the 12 glyphs was presented for 800 msec followed by a 200-msec ISI. Each 12-sec block was followed by a 4-sec rest period. A short interblock interval was used to make the transition between blocks relatively seamless, hence minimizing the likelihood that subjects would become aware of the two stimulus sets. Although this interval did not allow the fMRI signal to return completely to baseline, it was sufficient for our analysis because we focused on detecting a difference between conditions rather than estimating the hemodynamic response relative to a rest baseline. This single experimental run lasted approximately as long as the familiarization periods in typical visual statistical learning experiments (e.g., Fiser & Aslin, 2002), although triplets were only repeated half as many times due to the intermixed random blocks. Either before or after this run, half of the participants also completed one run of a different experiment (not reported here).

After the learning run, observers also completed a surprise familiarity test in the scanner. Similar to previous studies, this test involved 16 two-alternative forced-choice trials in which each triplet from the structured blocks was pitted against each foil sequence once (thus equating the frequencies of triplets and foils at test). Each glyph was presented in the same manner as during learning, with a 1-sec pause between the alternatives. Whether the triplet or the foil was presented first was randomized across trials. Observers used a keypress to indicate which alternative was more familiar.

After the test phase, observers completed a localizer run of the jiggle task containing four categories of stimuli in separate blocks: the glyphs from the learning run, line drawings of objects, grayscale faces, and four-letter English words (as used in Baker et al., 2007). There were eight 16-sec blocks of each category; stimuli were presented in the same manner as before but in random sequences.

Observers were asked five questions outside the scanner: (1) What do you think the experiment was about? (2) Did you use any particular strategy? (3) How do you think you did in the test phase? (4) Have you encountered an experiment like this before? (5) Did you notice any repeating patterns during the glyph jiggle task? These questions helped assess the implicitness of learning.

Image Acquisition

Neuroimaging data were collected on a 3T Siemens Trio scanner using a standard head coil. Functional data were acquired with a T2*-weighted gradient-echo, EPI sequence (TE = 25 msec; TR = 2000 msec; FA = 90°; matrix = 64 × 64) with 34 axial slices (3.5-mm isotropic voxels). For the learning run, 200 volumes were acquired; for the localizer run, 264 volumes were acquired. Two T1-weighted anatomical sequences were acquired for coregistration.

Preprocessing

The first four volumes of each functional run were discarded. Using Brain Voyager QX (Brain Innovation), data were then corrected for slice acquisition time, corrected for head motion, spatially smoothed (8-mm FWHM kernel), detrended, high-pass filtered with 128-sec period cutoff, normalized into Talairach space (Talairach & Tournoux, 1988), and interpolated to 3-mm isotropic voxels.

Data Analysis

To explore how neural responses differed for structured and random blocks, we used a general linear model treating block type as a fixed variable and subject as a random variable. The first block of each type was excluded from analysis because structure exists only insofar as the temporal patterns repeat across blocks; thus, differences between the first block of each type (before any triplets had been repeated) are not meaningful. Each block type was then entered as a separate regressor: A 12-sec boxcar function was defined for each of the remaining blocks of that type and convolved with a hemodynamic response function. As covariates of no interest, six regressors for each dimension of head movement were also included. This model estimated the contribution of each block type to the BOLD response in every voxel for each subject. The resulting beta values for the two conditions were compared across subjects using paired t tests. Voxels were judged to show a reliable difference for the contrast of structured versus random if the associated t value reached significance at $p < .001$ (two-tailed), and the voxel was part of a cluster of at least five contiguous voxels that all individually reached this significance level. Using the cluster-size threshold plug-in for BrainVoyager, which takes into account spatial smoothness (based on Forman et al., 1995), 10,000 Monte Carlo simulations revealed that the true corrected alpha associated with this significance level and cluster threshold is $p < .001$.

Three other analyses were conducted to further explore the difference between structured and random blocks. First, to examine whether statistical structure modulated activity in ventral visual cortex, we compared BOLD responses for structured versus random blocks within two *a priori* ROIs from the localizer: bilateral object-selective lateral occipital cortex (LOC) and left

word-selective ventral occipito-temporal cortex (VOTC). To define these regions in each subject, a multiple regression analysis similar to the one described above was used in which a different predictor was specified for each category. To localize the LOC, line drawings were contrasted against words and faces; to localize the VOTC, words were contrasted against faces and line drawings. In each region, the voxel with the greatest t value in an anatomically restricted search was selected as the center of a 4-mm sphere ROI if it reached at least $p < .001$. Responses were collapsed across peaks in bilateral dorsal and ventral aspects of LOC; however, the VOTC ROI was restricted to the left hemisphere. The LOC ROI could be defined in 15/16 observers, and the left VOTC ROI could be defined in 11/16 observers. This ROI-based approach was used in addition to the whole-brain analysis described above for two reasons: (1) the precise location of these functional ROIs is variable across subjects, and (2) these regions provide probes of category-specific visual processing in the ventral stream.

Second, to examine the speed of statistical learning, we created new predictors that estimated responses to pairs of blocks: For each condition, one regressor was defined for the hemodynamic response functions of blocks 1–2 (Epoch 1), blocks 3–4 (Epoch 2), etc., resulting in 12 regressors (structured, random × six epochs). Groupings of two blocks were selected as a compromise to minimize the noise associated with modeling a small number of time points while maximizing our resolution for detecting when the two conditions diverged.

Finally, we explored the relationship between the neural measure of statistical learning and the subsequent familiarity judgments. This relationship was assessed in two ways. First, we entered subsequent familiarity scores as a covariate in our model. In particular, in each voxel, the accuracy of familiarity judgments during the first half of the test phase was correlated with the magnitude of the difference between structured and random blocks; the covariate was obtained from the first half of the test because a significant effect was only observed during this phase (see below). Second, to examine whether neural evidence of statistical learning can be observed without familiarity, we obtained separate contrast maps for each subject (for the contrast of structured vs. random) and performed a one-sample t test of those subjects who performed at or below chance during the first half of the test (7 of 16). (Note that although deviation below chance could reflect some effect of learning, we could not assess the reliability of individual subject scores. Regardless, at a minimum, these subjects did not exhibit stronger familiarity with triplets than foils and thus would have been considered ''nonlearners'' in typical behavioral studies.) The strict threshold from our primary analysis was used for both of these analyses (cluster threshold corrected, $p < .001$).

RESULTS

Behavior

Cover Task

Performance on the jiggle task was very good overall (mean accuracy = 99.4%; mean response time = 446 msec) and did not differ between the structured and random blocks either in terms of accuracy, $t(13) = 1.38$, $p = .19$, $d = 0.37$, or response time, $t < 1$. This further confirms that the conditions were equated in all respects other than the presence of triplets in the structured blocks. Thus, any neural differences must reflect sensitivity to these regularities.

Familiarity Test

Preference for triplets in the familiarity test was weak (56%) and approached significance relative to chance, $t(15) = 1.69$, $p = .11$, $d = 0.42$. A more reliable effect was observed in the first half of the test phase (61%), $t(15) = 2.15$, $p < .05$, $d = 0.54$. The early part of the test phase may provide a better measure of familiarity of triplets from structured blocks because the same foils were repeated four times over the course of the test phase (to equate the frequency of triplets and foils at test) and thus would have become increasingly familiar. This fragile behavioral effect may be attributable to differences between the current study and previous studies, including that the structured blocks were interleaved with noise blocks (see Jungé, Scholl, & Chun, 2007); twice as many stimuli were used; and the jiggle task diverted attention away from the structure. Regardless, the robust neural effects described below provide an additional measure of statistical learning.

Debriefing

All participants were naive about the purpose of the study. In addition, 14/16 participants reported no awareness of any sequential patterns during learning even after being told about how the blocks were constructed. The two remaining participants claimed to have noticed some pairs (rather than triplets), but both participants performed below chance in the first half of the test phase. Thus, statistical learning operated without participants' awareness of the underlying structure between glyphs during the jiggle task (see also Turk-Browne et al., 2005; Saffran et al., 1997). In addition, observers' reported confidence about their performance during the familiarity test was low overall, and there was no obvious relationship between these reports and test accuracy: When asked to describe their performance, the only three observers with accuracies at or above 75% responded ''ok,'' ''guessing mostly,'' and ''terrible.''

Whole-Brain Analysis

To examine the neural correlates of statistical learning, we performed a voxel-wise contrast of structured versus random blocks. Several brain regions survived our statistical threshold (cluster threshold corrected, $p < .001$), showing enhanced responses to structured relative to random blocks (see Table 1). Of particular interest for assessing the relationship between statistical learning and other forms of learning and memory, responses to statistical structure were observed in the right striatum (caudate body; Figure 2A) and the right medial temporal lobe (hippocampus; Figure 2B) during learning. No regions exhibited the reverse pattern of random > structured.[1]

ROI Analysis

The effect of statistical structure in visual cortex was explored by comparing responses for structured versus random blocks in our LOC and VOTC ROIs (Figure 3B). Both regions exhibited stronger responses to structured than random blocks [LOC: $t(14) = 3.33$, $p = .005$, $d = 0.86$; VOTC: $t(10) = 2.44$, $p = .04$, $d = 0.73$].

Table 1. Structured > Random

Region	*Hemisphere*	*BA*	*x*	*y*	*z*	*Extent*	*Peak t*	*Epoch*
Limbic/Subcortical								
Hippocampus	R		27	−26	−4	13	5.84	3
Caudate	R		8	3	8	14	4.83	2
Parahippocampal gyrus	R	30	11	−42	−4	5	4.49	3
Frontal								
Medial/middle frontal gyrus	R	6	19	2	46	52	5.55	2
	R	8	11	38	36	16	5.35	*
Precentral gyrus	R	6	43	−3	34	60	5.25	2
	L	6	−33	−4	34	15	5.46	2
Insula	R	13	38	10	14	8	5.02	2
Middle frontal gyrus	R	9	28	22	32	8	4.73	5
Temporal								
Superior temporal gyrus	R	38	56	10	−19	17	5.59	2
Middle temporal gyrus	R	21	64	−8	−12	10	5.38	4
	R	37	46	−62	2	8	4.51	2
Fusiform gyrus	L	37	−47	−45	−10	10	4.94	2
Parietal								
Precuneus	R	7	19	−66	50	12	5.32	2
	L	19	−26	−75	32	25	6.00	*
Inferior parietal lobule	R	40	37	−53	33	27	5.72	4
Occipital								
Cuneus	R	19	26	−84	27	11	4.87	2
	L	18	−17	−77	16	10	5.32	2

Regions exhibiting stronger responses to structured versus random blocks ($n = 16$; $p < .001$; extent = five voxels; cluster threshold corrected, $p < .001$). BA, Brodmann's area; *x*/*y*/*z*, cluster center-of-gravity coordinates in Talairach space (Talairach & Tournoux, 1988); extent, number of suprathreshold voxels in the cluster; peak *t*, maximum statistic within cluster; epoch, point during run when evidence of learning first emerged ($p < .05$).

*Pairwise comparison between structured and random blocks did not reach significance in any individual epoch, but the difference was significant collapsed across epoch ($p < .001$).

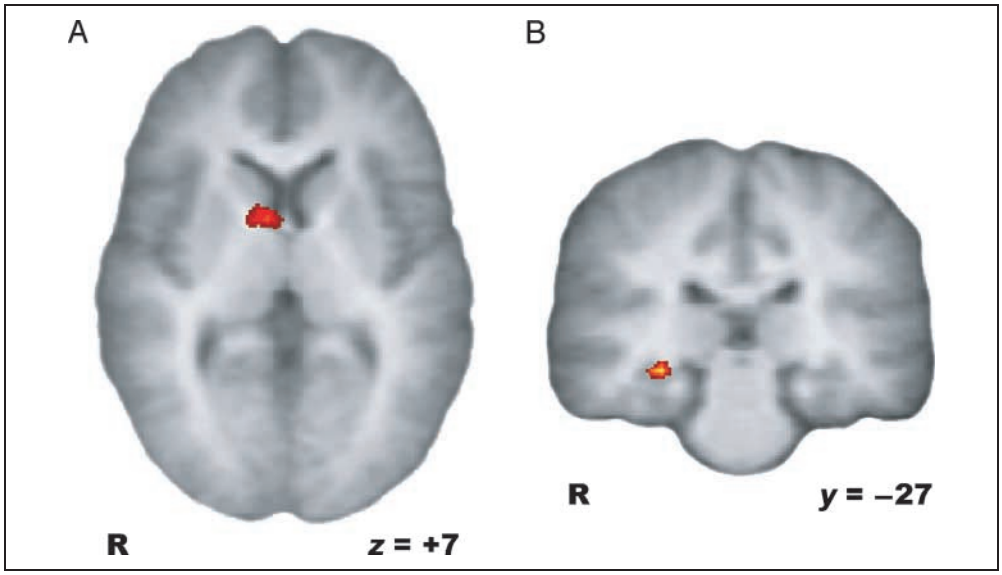

Figure 2. Learning regions. The right caudate (A) and the right hippocampus (B) exhibited stronger responses to structured versus random blocks. Activations are overlaid onto the average of all subjects structural T1 scans and smoothed for display purposes.

Epoch Analysis

To explore the efficiency of learning in the brain regions that showed whole-brain sensitivity to statistical structure, we assessed when the difference between structured and random blocks first emerged during learning. In particular, we defined epoch predictors using a moving window of two blocks: Epoch 1 modeled Blocks 1 and 2 of each condition, Epoch 2 modeled Blocks 3 and 4 of each condition, etc. Although the fact that these regions all exhibited a main effect of structured > random (collapsing over time) ensures that this difference would be observed eventually, this analysis nevertheless helps us determine how many triplet presentations are necessary to obtain such robust differences. Reassuringly, none of the regions exhibited a difference between structured and random blocks in Epoch 1 (ps > .15). However, significant differences (ps < .05) emerged beginning in Epoch 2. As can be seen in Table 1, some regions showed differences early (e.g., Epoch 2, in the

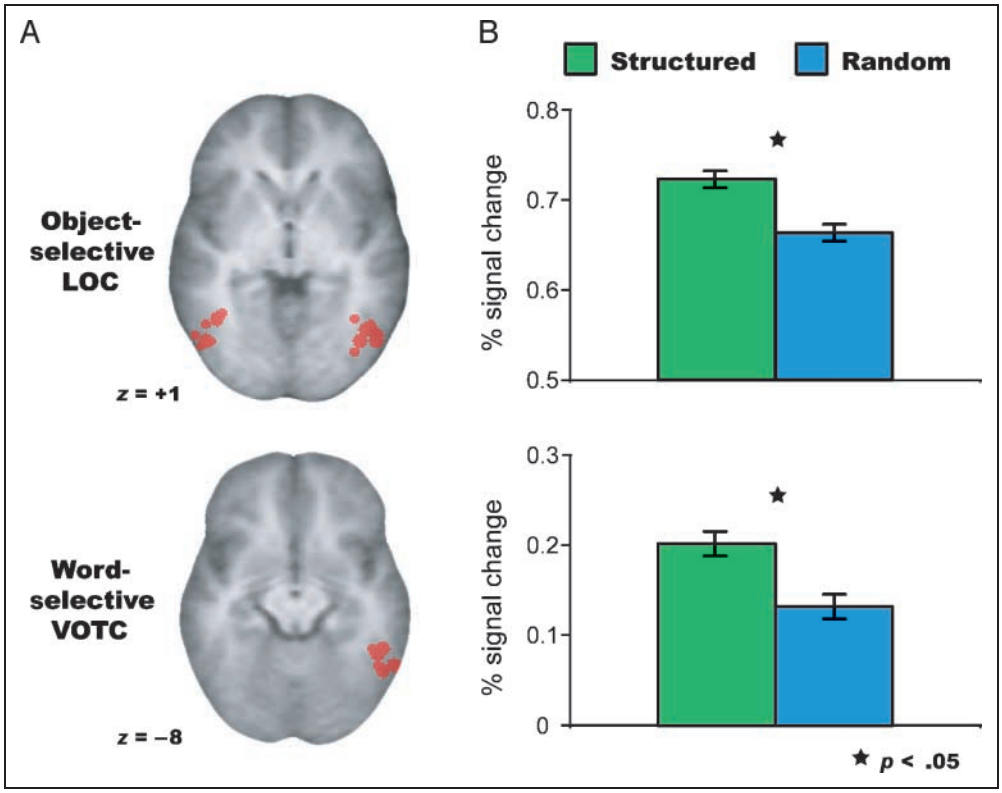

Figure 3. Visual regions. (A) Example of ventral object-selective and word-selective regions defined in each individual based on a separate localizer scan (some individuals visible as red spheres on average T1 scan). (B) BOLD responses for the structured and the random blocks extracted from each individual's ROIs. Error bars reflect within-subject standard errors of the differences between structured and random blocks in each region.

caudate and medial frontal gyrus) and some not until later (e.g., Epoch 4, in the middle temporal gyrus and inferior parietal lobule).

Relationship to Familiarity

To examine the relationship between neural and behavioral measures of learning, we correlated each observer's familiarity with the difference in activation between the structured and the random conditions in every voxel. The resulting r map was then thresholded (cluster threshold corrected, $p < .001$), revealing one region of the left frontal cortex (precentral gyrus/inferior frontal gyrus; Brodmann's area 6/9; the center of gravity of the cluster in Talairach coordinates: −57, 3, 27), where greater enhancement for structured versus random was associated with greater subsequent familiarity ($r = 0.81$, $p = .0001$).

To explore whether neural evidence of learning can be observed without familiarity, we also contrasted structured versus random blocks in the whole brain, excluding all subjects who performed above chance in the familiarity test. Notwithstanding the reduction in power associated with performing a random-effects analysis in seven observers, three regions exhibited a robust difference, including the original caudate region (cluster threshold corrected, $p < .001$; Table 2).

DISCUSSION

This study provides an initial exploration of the neural basis of visual statistical learning, using a design that has been employed in many behavioral studies. Our results have several implications:

Statistical Learning and Other Forms of Learning and Memory

The type of statistical learning explored in this study—learning of higher-order perceptual units—has often been treated as a distinct phenomenon (see Perruchet & Pacton, 2006). However, the extraction of regularities is an important component of other types of learning, including classification learning of associations between cues and outcomes (e.g., Knowlton, Squire, & Gluck, 1994), motor learning of sequences of manual responses (e.g., Nissen & Bullemer, 1987), and rule learning of generative grammars (e.g., Reber, 1967). Attempts have been made to relate these forms of learning to each other by appealing to the overlap in their neural bases—especially overlapping involvement of the striatal memory system (e.g., Poldrack et al., 2001; Knowlton, Mangels, & Squire, 1996). Although we cannot establish the necessity of the striatum for statistical learning from a correlational measure such as fMRI, our results suggest that the striatum may also be involved in statistical learning. In fact, the striatal region that activated to structured sequences in our study (the right caudate) is involved in many forms of implicit learning in humans (e.g., Seger & Cincotta, 2005; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Bischoff-Grethe, Martin, Mao, & Berns, 2001; Rauch et al., 1997) and animals (e.g., Winocur & Eskes, 1998; Packard, Hirsh, & White, 1989).

Our results also demonstrate that the medial-temporal lobe memory system may be involved in statistical learning. The involvement of the hippocampus may further help relate statistical learning to other forms of learning, including contextual learning (e.g., Chun & Phelps, 1999; cf. Manns & Squire, 2001), category learning (e.g., Cincotta & Seger, 2007), sequence learning (e.g., Ergorul & Eichenbaum, 2006; Schendan, Searl, Melrose, & Stern, 2003; Fortin, Agster, & Eichenbaum, 2002), and relational binding (e.g., Prince, Daselaar, & Cabeza, 2005; Mitchell, Johnson, Raye, & D'Esposito, 2000; Ryan, Althoff, Whitlow, & Cohen, 2000).

The involvement of both the caudate and the hippocampus in statistical learning raises the interesting possibility that parallel representations may be formed during statistical learning. One potentially relevant distinction is that striatal- and medial temporal lobe–mediated learning differ with respect to the flexibility of the resulting representations (see Johnson, van der Meer, & Redish, 2007): Learning involving the hippocampus may be more abstract and may generalize to new retrieval contexts, whereas learning involving the caudate may be specific and require exact replication of the encoding context for learning to be expressed. This distinction is apparent in animal studies of navigation in which the hippocampus is necessary for learning the spatial locations of rewards whereas the caudate is necessary for learning stimulus-response associations (e.g., Packard & McGaugh, 1996). Analogous effects have been observed in a human fMRI

Table 2. Unfamiliar Group

Region	*Hemisphere*	*BA*	*x*	*y*	*z*	*Extent*	*Peak t*
Medial frontal gyrus	R	10	4	52	7	5	13.96
Posterior cingulate cortex	R	30/31	31	−64	12	9	12.08
Caudate	R		10	3	16	7	9.60

Regions exhibiting stronger responses to structured versus random blocks in subjects that exhibited no subsequent familiarity ($n = 7$; $p < .001$; extent = five voxels).

study of navigation in which wayfinding—generating novel routes based on free exploration of a virtual environment—involved the hippocampus while following a familiar route involved the caudate (Hartley, Maguire, Spiers, & Burgess, 2003). The flexibility distinction is also present in human studies of visual associative learning, in which elderly participants with hippocampal atrophy fail transfer tests of learning whereas Parkinson's patients with basal ganglia dysfunction do not (Myers et al., 2003). Suggestively, in a different study of sequential visual statistical learning, we have found behavioral evidence of parallel abstract (order-invariant) and specific (order-specific) representations (Turk-Browne & Scholl, in press).

Statistical Learning and Perception

In addition to examining processes related to the detection and the extraction of perceptual regularities, we also assessed effects of statistical learning in visual cortex using a functional localizer. The two regions we identified are known to be highly selective for particular categories of visual stimuli—shapes/objects in the LOC (e.g., Malach et al., 1995) and characters/words in the VOTC (e.g., Baker et al., 2007)—and in terms of static visual stimulation, the structured and the random blocks in our study were perfectly equated. However, both regions responded more strongly in the presence of statistical structure, which existed only in terms of the distribution of glyphs within blocks. It is worth noting that this effect is surprising precisely because these are canonical ''object'' areas, where one might not expect sensitivity to ostensibly nonvisual distributed information. These results are consistent with the possibility that the human ventral stream may be selective for statistical information in perceptual input, akin to associative representations in monkey IT (e.g., Messinger, Squire, Zola, & Albright, 2001; Miyashita, 1993) and P-1 or P-2 learning in the MEM model of memory (Johnson & Hirst, 1993). Alternatively, greater responses to structured stimuli in perceptual areas may reflect attentional enhancement, possibly resulting from the downstream recognition of regularities. Any such effect must nevertheless be a consequence of statistical learning because the two block types only differed with respect to transitional probabilities. Moreover, because observers did not become consciously awareness of the structure, any preferential processing of the structured blocks was not voluntary.

Efficient Statistical Learning

Previous studies have been unable to examine the speed of statistical learning because they mostly relied on offline tests of learning after a preset (and largely arbitrary) amount of familiarization. No study, to our knowledge, has systematically varied the amount of familiarization to determine how quickly statistical learning can occur. One study used ERP as an on-line measure of auditory statistical learning (Abla et al., 2008). In this study, evidence of learning was examined over thirds of the familiarization stream, and one subset of the participants showed evidence of learning in the first third of familiarization. However, each third contained 40 presentations of every auditory regularity, and thus this finding provides a rather coarse estimate of efficiency. In contrast, we examined learning-related changes within the first dozen presentations of each visual regularity.

Our findings thus highlight the utility of fMRI as an incidental measure, able to provide an index of learning in progress with relatively high resolution. And surprisingly, neural evidence of statistical learning appeared very quickly: In some regions, reliable differences were observed in the second epoch, encompassing the third and the fourth blocks of each type, hence the third and the fourth presentation of each triplet. Because observers were naive about the length of our regularities, this may be the minimal number of blocks necessary for each triplet to appear variably with respect to the other triplets. On the other hand, learning-related neural changes may have begun as early as the second presentation but might not have been reliably detected until the second epoch due to power limitations (e.g., number of participants, limits of fMRI). Nevertheless, robust evidence of learning was observed very early, although participants were performing an orthogonal task and were largely unaware of the structure.

Varieties of Statistical Learning: Implicit Neural Responses versus Explicit Familiarity Judgments

We observed clear neural evidence of statistical learning that was substantially more robust than the conventional behavioral familiarity measure. This greater sensitivity of the neural measure is important because it highlights that learning was largely implicit. Specifically, participants reported no awareness of the structure during debriefing, and the neural evidence of learning was accompanied by weak explicit familiarity. One region of left frontal cortex did strongly correlate with familiarity across observers but did not overlap with the regions showing an overall effect of statistical structure.

Neural measures are especially useful in this context because they can reveal evidence of learning that has not or will not be manifested in behavior (e.g., McNealy et al., 2006; Landau, Schumacher, Garavan, Druzgal, & D'Esposito, 2004). In our case, a subset of observers who exhibited no subsequent familiarity with the regularities still showed a robust difference between structured and random blocks in the striatum. Because the two block types were completely balanced for the novelty and block positions of individual glyphs, this difference must reflect learning of the transitional probabilities in the structured blocks. Typically, these observers would be categorized as ''nonlearners,'' but this

may be an unfair characterization given our results. Instead, the typical behavioral identification of learners versus nonlearners may reflect a stable individual difference either in the ''rate'' of learning (with explicit familiarity reflecting a later stage of the learning process) or in the ''mode'' of learning: Some observers may learn more implicitly (resulting in strong activation of the caudate), whereas others may learn in a manner that is more conducive to explicit familiarity (resulting in activation of regions supporting such processing).

Similar individual differences in statistical learning have been reported previously. For example, in the ERP study discussed earlier (Abla et al., 2008), participants were divided into ''high,'' ''middle,'' and ''low learners'' based on subsequent familiarity. Although high and middle learners differed in terms of familiarity (by definition), similarly robust learning-related ERP changes were observed in both groups. However, these changes were observed only in the first third of familiarization for high learners and only in the final third of familiarization for middle learners, again demonstrating that lower familiarity does not necessarily imply less learning per se. Our results provide even stronger evidence for this claim because robust neural evidence of learning was observed in participants who expressed no familiarity whatsoever. In sum, implicit measures of online learning can be both more sensitive and provide richer detail about the learning process itself than explicit measures taken later.

Conclusions

Collectively, these results emphasize both the power of statistical learning and its integration with other cognitive processes. The neural responses observed here provide clear evidence that statistical learning can occur implicitly and quickly, and that neuroimaging may be an especially sensitive technique for exploring such processing. The particular brain regions that reflected this learning help tie statistical learning to other forms of learning and perceptual processing.

Acknowledgments

We thank Malte Alf, Riana Betzler, Harlan Fichtenholtz, Julie Golomb, Christian Luhmann, and Greg McCarthy for helpful conversations. Portions of this research were presented at the 2008 meeting of the *Vision Sciences Society*. Supported by NIH grants AG09253 (M. K. J.), EY014193 and P30 EY000785 (M. M. C.), and NSERC PGS-D (N. T. B.).

Reprint requests should be sent to Nicholas B. Turk-Browne, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06511, or via e-mail: nicholas.turk-browne@yale.edu.

Note

1. The repetition of visual information typically causes the attenuation of evoked responses in selective ventral occipital and temporal regions (for a review, see Grill-Spector, Henson, & Martin, 2006). The failure to observe attenuated responses for repeated sequences (i.e., random > structured) may thus be related to the fact that individual glyphs in both structured and random blocks were repeated several times such that visual-evoked responses were fully adapted (see Reber, Gitelman, Parrish, & Mesulam, 2005). Nevertheless, the interplay between repetition attenuation for items versus associations deserves more consideration (cf. Köhler, Danckert, Gati, & Menon, 2005). Moreover, despite the lack of a decrease for structured versus random blocks overall, learning-related decreases may still have occurred for structured blocks later versus earlier during learning. Given the challenges discussed earlier in making comparisons between conditions occurring at different points in an fMRI session, we chose a contemporaneous baseline task. As a result, it remains possible that both structured and random responses may have decreased over time, with a more precipitous decline in the random condition.

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20,* 952–964.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 9087–9092.

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

Bischoff-Grethe, A., Martin, M., Mao, H., & Berns, G. S. (2001). The context of uncertainty modulates the subcortical response to predictability. *Journal of Cognitive Neuroscience, 13,* 986–993.

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

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

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience, 2,* 844–847.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience, 19,* 249–265.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24–39.

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

Ergorul, C., & Eichenbaum, H. (2006). Essential role of the hippocampal formation in rapid learning of higher-order sequential associations. *Journal of Neuroscience, 26,* 4111–4117.

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.
Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review, 14,* 173–178.
Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance Medicine, 33,* 636–647.
Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience, 5,* 458–462.
Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7,* 497–510.
Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10,* 14–23.
Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled distinct neural bases of route following and wayfinding in humans. *Neuron, 37,* 877–888.
Johnson, A., van der Meer, M. A., & Redish, A. D. (2007). Integrating hippocampus and striatum in decision-making. *Current Opinion in Neurobiology, 17,* 692–697.
Johnson, M. K., & Hirst, W. (1993). MEM: Memory subsystems as processes. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 241–286). East Sussex, England: Erlbaum.
Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition, 15,* 1–11.
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.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399–1402.
Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1,* 106–120.
Köhler, S., Danckert, S., Gati, J. S., & Menon, R. S. (2005). Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: A comparison based on event-related fMRI. *Hippocampus, 15,* 763–774.
Landau, S. M., Schumacher, E. H., Garavan, H., Druzgal, T. J., & D'Esposito, M. (2004). A functional MRI study of the influence of practice on component processes of working memory. *Neuroimage, 22,* 211–221.
Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience, 16,* 427–438.
Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 8135–8139.
Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus, 11,* 776–782.
McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience, 26,* 7629–7639.
Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12239–12244.
Mitchell, K. J., Johnson, M. K., Raye, C. L., & D'Esposito, M. (2000). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive Brain Research, 10,* 197–206.
Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience, 16,* 245–263.
Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15,* 185–193.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.
Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32.
Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 2745–2750.
Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience, 9,* 1465–1472.
Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory, 65,* 65–72.
Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 80–96.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10,* 233–238.
Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550.
Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural correlates of relational memory: Successful encoding and retrieval of semantic and perceptual associations. *Journal of Neuroscience, 25,* 1203–1210.
Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., et al. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping, 5,* 124–132.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.
Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2005). Priming effects in the fusiform gyrus: Changes in neural activity beyond the second presentation. *Cerebral Cortex, 15,* 787–795.

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science, 11,* 454–461.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.
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,* 101–105.
Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the role of the medial-temporal lobe in implicit and explicit sequence learning. *Neuron, 37,* 1013–1025.
Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25,* 2941–2951.
Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2002). Neural correlates of artificial grammar learning. *Neuroimage, 17,* 1306–1314.
Talairach, J., & Tournoux, P. (1988). *A co-planar stereotactic atlas of the human brain*. New York: Thieme.
Toni, I., Krams, M., Turner, R., & Passingham, R. E. (1998). The time course of changes during motor sequence learning: A whole-brain fMRI study. *Neuroimage, 8,* 50–61.
Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 399–407.
Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.
Turk-Browne, N. B., & Scholl, B. J. (in press). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*.
Winocur, G., & Eskes, G. (1998). Prefrontal cortex and caudate nucleus in conditional associative learning: Dissociated effects of selective brain lesions in rats. *Behavioral Neuroscience, 112,* 89–101.
Young, R. K. (1968). Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 122–148). Englewood Cliffs, NJ: Prentice-Hall.