Statistical Learning and Its Consequences

Nicholas B. Turk-Browne

Abstract Statistical learning refers to an unconscious cognitive process in which repeated patterns, or regularities, are extracted from the sensory environment. In this chapter, I describe what is currently known about statistical learning. First, I classify types of regularities that exist in the visual environment. Second, I introduce a family of experimental paradigms that have been used to study statistical learning in the laboratory. Third, I review a series of behavioral and functional neuroimaging studies that seek to uncover the underlying nature of statistical learning. Finally, I consider ways in which statistical learning may be important for perception, attention, and visual search. The goals of this chapter are thus to highlight the prevalence of regularities, to explain how they are extracted by the mind and brain, and to suggest that the resulting knowledge has widespread consequences for other aspects of cognition.

Keywords Regularities · Memory systems · Perception · Selective attention · Generalization · fMRI

Introduction

Human behavior is often geared towards one object at a time, as in picking up a coffee mug, recognizing a friend's face, or noticing a car's age. This fact is even more apparent in visual search, where we typically seek one target object among other distracting objects: looking for my coffee mug among many others in the office lounge; trying to track down a particular friend at a cocktail party; or, searching for my car in an airport parking garage. How we succeed (and fail) in these kinds of searches is the topic of the 59th Nebraska Symposium, including critical factors such as attention, memory, reward, and real-world complexities. The purpose of this chapter is to highlight another important factor in visual search, 'statistical learning'.

N. B. Turk-Browne (✉)
Department of Psychology, Princeton University, Green Hall,
Princeton, NJ 08540, USA
e-mail: ntb@princeton.edu

M. D. Dodd, J. H. Flowers (eds.), *The Influence of Attention, Learning, and Motivation on Visual Search,* Nebraska Symposium on Motivation,
DOI 10.1007/978-1-4614-4794-8_6,

Although the goal of visual search is to find a target object, we rarely need to start from scratch. Rather, we can use knowledge about when and where this object appears relative to other objects to find what we're looking for. In the searches above, for example, I may know from prior experience that my coffee mug sits on top of a shelf rather than floating in air; that my friend hangs out with certain people who may also be at the party; and, that I tend to park near the elevator in parking garages. Indeed, we repeatedly come across the same people, places, and things, and over time they tend to appear in similar spatial configurations and temporal sequences. Statistical learning is an unconscious process by which we extract these patterns (or 'regularities') in how objects appear relative to each other in the visual environment.

Statistical Regularities in the Visual Environment

Regularities are aspects of the environment that repeat over time, such as the fact that beaches tend to look the same, that football players tend to appear on football fields, or that my office is across the street from a pharmacy and an ice cream shop. These regularities can be roughly classified along two dimensions: the timescale of learning (the interval over which learning happens), and the domain of knowledge (what kind of information is learned). Three varieties of regularities that differ on these dimensions are reviewed below: (1) regularities that have molded the visual system over evolutionary time and extensive training to basic physical properties of the environment ('physical regularities'), (2) regularities that are acquired throughout the lifespan about kinds of objects grouped as visual concepts/categories ('semantic regularities'), and (3) regularities that are learned in minutes about happenstance relationships between particular features and objects ('token regularities'). The boundaries between these types are inherently fuzzy, and the separation below is not intended to reify strict distinctions. Rather, this separation is used rhetorically to provide intuitions about the prevalence of regularities in many aspects of perception.

Physical Regularities

The natural environment has been stable for a long time, and so the range of input received by the visual system is constrained. For example, natural landscapes have horizontal but not vertical horizons, and natural light comes from above but not below. Over phylogenetic time, our brains have adapted to such regularities: if a large set of natural images is decomposed into independent basis functions with a constraint on efficient coding, the resulting components match the receptive field properties of V1 neurons (Olshausen and Field 1996). In other words, the visual system has been tuned during evolution to the statistics of the natural world (Simoncelli and Olshausen 2001).

Beyond regularities in the general appearance of natural images, there are also physical constraints on how objects in the world interact. For example, objects move continuously through space and time, objects only rest on horizontal surfaces unless affixed, and objects 'touch' their shadows only where they rest on the obscured surface. These and other constraints may be built into the visual system, as evidenced by studies of infant cognition. For example, 3–4 month old infants know that objects continue to exist after being occluded and cannot pass through each other (Baillargeon 1987). While present early on, knowledge of physical regularities becomes more sophisticated throughout development, growing from knowledge about simple occlusion to, for example, knowledge about containment and transparency (Baillargeon 2008).

Along with changes over phylogenetic time, extensive experience with specific input can alter the visual system over ontogenetic time. For example, repeated subliminal exposure to one direction of motion improves the detection threshold for that direction but not other directions (Watanabe et al. 2001). Such 'perceptual learning' reflects long-lasting changes in visual cortex, including enlarged and refined representations of trained stimuli (Goldstone 1998; Fahle and Poggio 2002; Sasaki et al. 2010). This kind of learning may supplement hard-wired general assumptions about the visual environment (as described above) by further tuning the visual system to the natural and artificial environments that we inhabit during our lifetime.

Semantic Regularities

In addition to physical properties and object interactions, regularities exist in how *types* of objects appear in the visual environment. For example, regardless of color and shape, all fire hydrants appear on sidewalks not on top of mailboxes, and all toilets appear in bathrooms and not kitchens. The visual system relies on these canonical locations and scene contexts for object recognition, such that objects violating semantic regularities are recognized more poorly (Palmer 1975; Biederman et al. 1982; Davenport and Potter 2004; cf. Hollingworth and Henderson 1998). In addition, there are semantic regularities in terms of whether and how objects can change over time. For example, stoplights can change color but stop signs can't, and humans can move locations but not shrink in size (at least not quickly). These kinds of regularities influence visual awareness: improbable object changes have a higher incidence of change blindness (Beck et al. 2004).

The distinction between semantic and physical regularities is not rigid. For example, images representing the same scene category (e.g., a beach) have similar global physical properties. Despite differences in the local details of particular exemplars (e.g., the number of bathers, color of sand, presence of beach chairs, etc.), such properties allow us to quickly and effortlessly categorize novel scenes (Oliva and Torralba 2006). However, the objects belonging to a semantic regularity can *but need not be* especially physically similar. For example, the exemplars of many categories (e.g., buildings, trees, dogs) are physically heterogeneous. More generally, while physical properties can denote particular semantic classes, this is just one type of linking cue: objects may share semantic relations in spite of physical dissimilarity because, for example, they accomplish a similar function (e.g., microwave and BBQ grill) or are required components of a broader concept (e.g., the equipment, players, and field in football).

Token Regularities

In addition to regularities in what semantic types of objects co-occur, the visual environment is littered with regularities between object *tokens*. For example, when learning a new navigation route, the sequence of landmarks that one comes across is highly regular (e.g., a religious billboard, then a fast food restaurant, then an open field, etc.). Such regularities can be spatial in addition to temporal. For example, beyond physical and semantic constraints, there are regularities in the configuration of objects in each kitchen that are unique with respect to all other kitchens. Finally, some token regularities are not inherently spatial or temporal, such as sets of faces that tend to co-occur (e.g., in groups of friends).

There is no necessary semantic relationship between objects in token regularities (e.g., a billboard and a restaurant, two facial identities), and indeed many such regularities are composed of conceptually distant objects (e.g., the curtains, electronics, and art in a room). More importantly, semantic classes often provide too coarse a grain of description (e.g., moving around my living room in the dark or finding a particular utensil in the kitchen does not benefit much from my general knowledge of living rooms or kitchens). Token regularities also differ from the kinds of physical regularities described earlier that persist over very long timescales: particular sequences and configurations of objects are happenstance and can change frequently. Moreover, I will emphasize the *relational* nature of token regularities—that regularities exist in the statistical relationships (e.g., conditional probabilities) between two or more objects—whereas perceptual learning of physical regularities is often viewed as imprinting of discrete stimuli (Goldstone 1998). In sum, statistical learning about token regularities occurs over faster timescales (e.g., when moving to a new city, or using a new computer operating system) and requires tracking particular object exemplars rather than semantic classes or physical properties per se.

Studying Statistical Learning

Unlike physical and semantic regularities, which are built into the visual system or learned over the course of development, token regularities can be introduced and tested in a laboratory setting. In addition to introducing new regularities, experiments about statistical learning typically use novel objects without preexisting semantic associations. Thus, subjects begin *de novo* when entering these experiments, with minimal prior knowledge and biases relevant to the regularities (or to the objects from which they are generated). This experimental control allows us to characterize the operation of statistical learning precisely. Given the space of possible regularities in the real world, statistical learning is generally studied in an artificial environment containing a small number of relatively simple regularities. Whether these environments provide a suitable proxy for how statistical learning operates in more naturalistic contexts is an important and open question. Later, I will consider evidence that statistical learning can handle some of the complexities of natural environments.

Fig. 1 Auditory statistical learning. Subjects are briefly exposed to a structured speech stream, and subsequently express greater familiarity with words than part-words and non-words (see text for details). (Based on Saffran et al. 1996a)

Where It All Started: Auditory Statistical Learning

This chapter is predominantly focused on *visual* statistical learning. But the impetus for current research on visual statistical learning (including the author's own initial interest), comes from the study of auditory statistical learning; in particular, from studies of how auditory statistical learning may be important for word learning and segmentation in speech streams, and for language acquisition more generally (Saffran et al. 1996a,b).

In a typical study of this type (Fig. 1; Saffran et al. 1996a), subjects (infants, kids, or adults) are exposed to a brief speech stream composed of syllables. Unbeknownst to subjects, the stream of syllables has been constructed to contain statistical regularities in terms of which syllables follow each other. In particular, the 12 total syllables (e.g., tu, pi, ro, bi, da, ku, go, la, bu, pa, do, ti) have been assigned to four trisyllabic 'words' (e.g., tupiro, bidaku, golabu, padoti), such that the first syllable is always followed by the second, and the second always by the third (transitional probabilities of 1.0). What occurs after the third syllable of each word is any of the first syllables from the three other words (back-to-back word repetitions are prevented; transitional probabilities of 0.33). Insofar as subjects pick up on the statistical discrepancy between within- and between-word transitions, they can use the transitional probabilities to segment the words. Critically, other than the differential transitional probabilities, there are no prosodic or other cues that would indicate word boundaries.

After only 2 min of exposure, 8-month-old infants expressed knowledge of the words by dishabituating to new words that were composed of the same syllables, but in combinations with lower (or zero) transitional probabilities based on what they had heard (Saffran et al. 1996a). While not discussed further here, these results have been extended significantly to incorporate, among other things, prosody (e.g., Thiessen and Saffran 2003), word meaning (e.g., Graf Estes et al. 2007), and multiple languages (Gebhart et al. 2009).

Temporal Visual Statistical Learning

As highlighted earlier, statistical regularities are ubiquitous in the visual environment. Regularities exist in both how objects are arrayed in space and how they appear over time. Tasks have been developed to study statistical learning in both of these dimensions. Visual statistical learning about temporal regularities is considered first, because it is studied in a similar way to auditory statistical learning. Temporal information in vision arises in one of at least two ways:

First, the world is dynamic and objects move over time. The way that objects move and change over time contains regularities, such as the trajectory of a tennis ball, the different views of a car as it passes, or the sequence of movements underlying an action (Baldwin et al. 2008). These kinds of temporal regularities have an intrinsic structure or order, and may underlie our representations of events (Avrahami and Kareev 1994; Zacks and Tversky 2001).

Second, the way that we sample the world (both static and dynamic parts) guarantees that visual input will be distributed over time. Specifically, we acquire information about one part of the visual environment at a time through eye movements and covert shifts of attention. We must continually sample the environment due to the capacity limitation of visual short-term memory—i.e., the small number of items that can be held simultaneously in memory and the brief durations of such representations (e.g., Henderson and Hollingworth 2003; Zhang and Luck 2009). Objects captured by each eye fixation or attentional shift receive enhanced processing, such that scanning creates a parade of objects through the visual system. Temporal regularities of this type may not be inherently ordered or structured, but can exist in terms of which objects appear in a broader context (e.g., the set of objects in one room), or can be driven by the likelihood that two or more objects will be fixated sequentially (e.g., because of close spatial proximity or similar salience). Thus, temporal regularities are prevalent in vision as well as in audition, due to the existence of inherently temporal events and actions, and to our serial sampling of spatial environments.

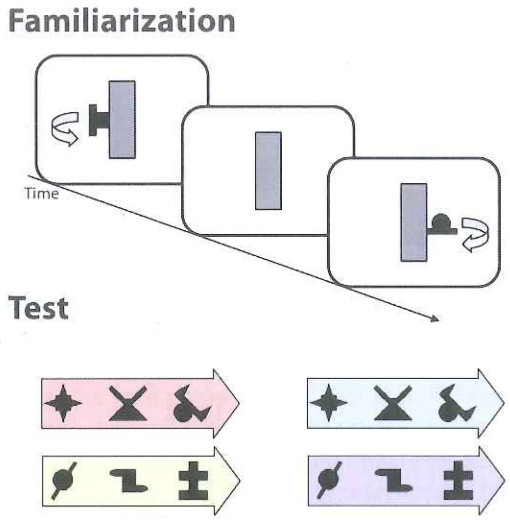

Fig. 2 Temporal visual statistical learning. Subjects are briefly exposed to a structured stream of shapes, and subsequently express greater familiarity with triplets than foils (see text for details). (Based on Fiser and Aslin 2002)

Studies of temporal visual statistical learning often rely on an experimental design adapted from the original auditory statistical learning experiments (Fig. 2; Fiser and Aslin 2002; see also Olson and Chun 2001). A set of novel nonsense shapes (e.g., A, B, C, D, E, F, G, H, I, J, K, L) is divided without replacement into temporal pairs or triplets (e.g., ABC, DEF, GHI, JKL). During an initial phase, subjects are exposed to a continuous stream constructed from these triplets, with shapes appearing one at a time (e.g., DEFJKLABCDEFGHI. . .). Critically, subjects are not oriented to the presence of triplets. The shapes can either appear as the only thing on the screen, or cycle back and forth behind an occluder changing identity each time when occluded.

After several minutes of exposure, subjects are then given a surprise familiarity test. On each trial they are presented with two three-item sequences: (1) a triplet from familiarization, and (2) a foil generated from the same shapes but rearranged into new groupings (e.g., AEI, DHL, GKC, JBF). Each triplet is tested against each foil, to ensure equal frequency of the alternatives at test. Thus, triplets can be discriminated from foils only based on the higher transitional probabilities within triplets vs. within foils. Subjects perform very well in this task, despite the fact that learning is incidental and that subjects often express low confidence in their test judgments. Such findings have been used to suggest that statistical learning happens automatically as a result of mere exposure to regularities (Saffran et al. 1999; Fiser and Aslin 2002).

Spatial Visual Statistical Learning

Temporal regularities derive partly from repeated sampling of structured spatial environments. Indeed, objects and their parts do not appear in random locations in scenes, but rather in predictable locations based on the locations of other objects and

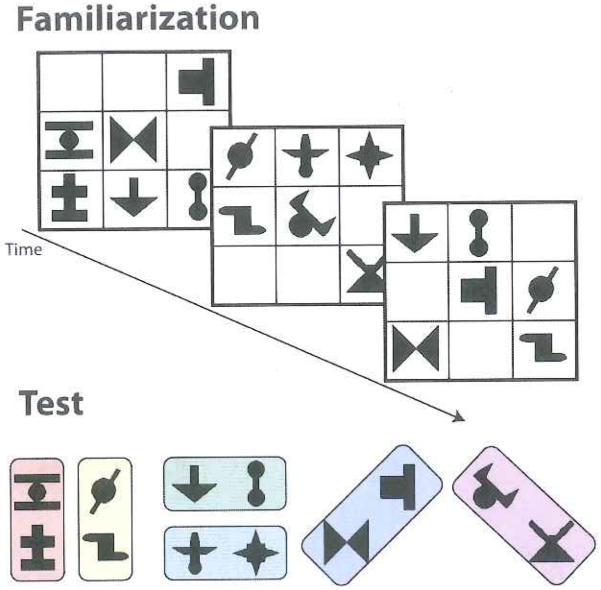

Fig. 3 Spatial visual statistical learning. Subjects are briefly exposed to a series of grids containing spatial pairs, and subsequently express greater familiarity with pairs than foils (see text for details). (Based on Fiser and Aslin 2001)

parts (Biederman et al. 1982). These spatial regularities are partly due to physical constraints (sofas cannot float in the sky) and partly due to semantic knowledge (hood fans appear above rather than below stoves). However, additional regularities exist in the particular configuration of parts and objects. Sensitivity to these regularities may be important for learning about the composition of novel objects and about the layout of novel environments. Indeed, space is the dominant dimension in the visual modality (as time is in the auditory modality; Kubovy 1988), and thus it is important to characterize how statistical learning operates over spatial regularities.

Studies of spatial visual statistical learning employ a design that is somewhat different from temporal statistical learning studies. However, at its core, this design replicates the essential property of temporal statistical learning experiments: that regularities can only be segmented on the basis of statistics. In other words, much like the continuous temporal stream from which shape triplets must be extracted, our visual system is confronted with an undifferentiated spatial image and must segment the meaningful chunks based solely on spatial probabilities.

In such studies (Fig. 3; Fiser and Aslin 2001; see also Chun and Jiang 1999), subjects are presented with simplified visual scenes generated from a 3 × 3 grid. Six shapes are shown in each grid and—unbeknownst to subjects—the shapes appear in spatial pairs. Specifically, a set of 12 shapes (e.g., A, B, C, D, E, F, G, H, I, J, K, L) is randomly assigned without replacement to six pairs (e.g., AB, CD, EF, GH, IJ, KL). Two pairs are assigned to each of three orientation types: horizontal (e.g., AB, CD), vertical (e.g., EF, GH), and diagonal (e.g., IJ, KL). Each scene is generated by selecting one pair of each orientation (e.g., AB, EF, IJ) and placing them on the grid so that all shapes are abutted by at least one shape beyond the paired shape. With these constraints, a total of 144 grids can be generated. Subjects are briefly presented with each scene and the entire stream lasts 5 min. Critically, to learn the pairs, subjects need to extract the joint probabilities between adjacent shapes. To measure learning, subjects then complete a familiarity test in which they are repeatedly presented with a pair and a foil (e.g., AD, CB, EH, GF, IL, KJ) and must choose which alternative is most familiar. Performance at discriminating pairs from foils is excellent in this task.

A modified design provides a test of *which* statistics subjects can extract. In the basic task, joint and conditional probabilities are confounded (pairs have higher joint and conditional probabilities than foils): i.e., $P(AB)=0.5>P(AD)=0$ and $P(B|A)=1.0>P(D|A)=0$. To examine whether subjects could pick up conditional probabilities in isolation, Fiser and Aslin (2001) doubled the frequency of a subset of the pairs (e.g., AB, IJ) such that the joint probability of two adjacent shapes that crossed pairs (e.g., BJ) was equal to the joint probability of infrequent original pairs (e.g., CD). Importantly, these two pair types differed in conditional probability: $P(D|C)=1>P(J|B)\sim 0.5$. Subjects were still able to successfully discriminate these pairs, suggesting that they had extracted conditional probabilities. In contrast to joint probabilities, conditional probabilities may be especially important for prediction (see 'Anticipation' section).

Aside: Other Related Kinds of Learning

Visual statistical learning as a cognitive process may occur in many task contexts. Most prominently, a form of visual statistical learning may underlie contextual cueing during visual search (Chun and Jiang 1998). In such tasks, subjects are presented with visual search arrays composed of a T target and L distractors. Critically, several configurations of targets and distractors are repeated during the experiment, and search performance gets faster and faster as a result of learning. This task provides an elegant online measure of learning, allowing the timecourse of learning to be assessed, unlike most statistical learning tasks. However, the spatial regularities in contextual cueing are qualitatively different from statistical learning: subjects learn configurations of locations *per se*, rather than configurations of object identities. Thus, contextual cueing may help in locating a target during visual search given the known locations of other objects (cf. Kunar et al. 2007), while spatial statistical learning results in stimulus-specific knowledge of relative locations (e.g., that A is above B). Variations on contextual cueing in which the identities of distractors predict the location and/or identity of the target (Chun and Jiang 1999; Endo and Takeda 2004) provide a bridge to the spatial statistical learning literature. There are many other related forms of learning, including in the serial reaction time task where sequences of spatial locations (rather than object identities) can be readily learned (e.g., Mayr 1996).

The Nature of Statistical Learning

Statistical regularities are everywhere in the world, and evidence of statistical learning is almost equally ubiquitous. Statistical learning occurs in every sensory modality in which it has been tested (e.g., Conway and Christiansen 2005), in time and in space (e.g., Fiser and Aslin 2001, 2002), for many types of stimuli (e.g., Saffran et al. 1999; Baldwin et al. 2008), and in many subject populations (e.g., Kirkham et al. 2002; Toro and Trobalón 2005). Statistical learning seems to be a powerful and fundamental part of cognition... but how does it work? Over the past seven years, my collaborators and I have attempted to uncover the nature of statistical learning. Here I review findings from this research program as an update on what is known about statistical learning.

When Does Learning Take Place?

An important challenge for statistical learning is that the huge set of regularities we experience is a small subset of all possible regularities that could exist. For example, consider walking through your home airport: you may repeatedly encounter the same restaurants, ticket agents, and gates, and they can even appear in fixed temporal orders and spatial layouts; but on any given trip, you encounter many other objects, such as random tourists, food stands, bestsellers in the bookstore, etc., and these objects may be gone by your next trip. How do we extract the meaningful and stable relationships between certain objects, while discounting transient intrusions from other objects? The problem is one of 'combinatorial explosion': to learn about which relationships are regular, one must in principle represent the co-occurrence of all possible groupings of objects. In other words, how can statistical learning determine *a priori* which relationships will reappear in the future? Indeed, if such advance knowledge were possible, learning would have already taken place! This is especially problematic for statistical learning since it occurs in an unsupervised manner (Fiser and Aslin 2002), without feedback about when and what to learn.

In a series of behavioral studies (Turk-Browne et al. 2005), we explored this issue by examining the automaticity of visual statistical learning. Namely, we asked whether statistical learning occurs whenever the visual system is confronted with regularities, or whether selective attention can determine which regularities are learned. The latter possibility would place an important constraint on when statistical learning takes place. In our studies, subjects were presented with a temporal stream of nonsense shapes. This stream was itself composed of two separate streams that had been interleaved (Fig. 4). Each stream appeared in a different color, and contained a unique set of shapes. Unbeknownst to subjects, each of the two colored streams was generated using triplets of shapes as in the standard temporal visual statistical learning task.

Critically, subjects were instructed to attend to one color of shapes (e.g., green), and to perform a one-back task only for shapes appearing in that color (i.e., detecting whether the current green shape was the same as the last green shape). Shapes were presented one at a time, and were thus all attended spatially. However, only the attended shapes were task-relevant. Insofar as learning occurs by mere exposure to regularities (e.g., Saffran et al. 1999), regularities in both streams may be learned. If selective attention instead gates statistical learning, we would expect better learning of the attended shapes.

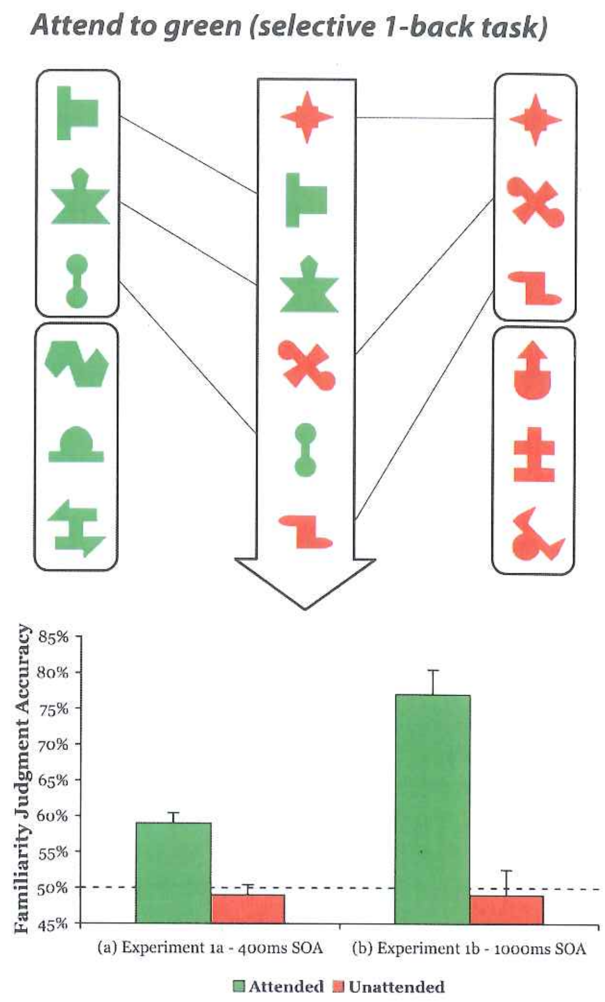

Fig. 4 Selective attention. Familiarization stream constructed from two sets of regularities, one attended and the other unattended. Statistical learning only occurs for attended regularities. (Based on Turk-Browne et al. 2005)

Across five experiments, we not only found better learning of the attended regularities, but no learning whatsoever of the unattended regularities. This was even true when we used an implicit response time (RT) measure that may be more sensitive to unattended learning. Thus, selective attention determines the input to statistical learning. These results were recently used as a case study of how to prove null hypotheses (Gallistel 2009)—in our case, the null hypothesis was that *no* learning occurs without attention (confirmed by way of Bayesian analysis).

While statistical learning of object relationships occurred only when the objects were task-relevant, this learning happened without conscious awareness. Indeed, statistical learning was robust despite the fact: (1) that subjects were not informed about the presence of regularities, (2) that they performed a distracting cover task (one-back) during familiarization, (3) that the shapes were presented quickly, (4) that regularities from the two streams were interleaved, adding noise to the transitional probabilities, and (5) that learning was evident in an implicit RT measure. Moreover, during careful debriefing in the RT experiment, no subjects expressed awareness of the structure in the displays. These findings suggest that statistical learning is and is not automatic: selective attention to objects is required for their relationships to be learned, but once this input has been selected, learning takes place without conscious intent or effort.

Does Task-Relevance Guarantee Learning?

The previous section described evidence that statistical learning is constrained by top-down selective attention, i.e. that statistical learning only occurs for task-relevant objects. However, the simple working memory task used in those experiments is only one of a large number of cognitive tasks that we routinely engage in. Does statistical learning take place whenever objects are task-relevant, or does the nature of the task matter? In the latter case, finding that some tasks are better or worse for statistical learning may help uncover the component processes at work during statistical learning.

In a recent behavioral study (Zhao et al. 2011), we examined one particular kind of task that bears an interesting resemblance to statistical learning: statistical summary perception. The ability to perceive summary statistics has received much consideration recently (e.g., Ariely 2001; Chong and Treisman 2003; Alvarez and Oliva 2008). In statistical summary perception tasks, subjects are presented with a set of objects, and are instructed to make subsequent judgments about some statistical property of the set (e.g., mean size). Critically, subjects can often extract summary statistics from a set of objects without being able to identity the constituent members of the set (e.g., Ariely 2001; Alvarez and Oliva 2008).

On the surface, statistical learning and statistical summary perception are quite different: statistical learning involves extracting regularities over repeated experience, while statistical summary perception involves extracting statistics from a single display; and statistical learning involves acquiring stimulus-specific relationships (i.e. that this particular object co-occurs with another particular object), while statistical summary perception (by definition) involves representing the general properties of a collection. Despite these surface differences, however, these two processes are both inherently statistical: they involve aggregating a sample, and distilling this sample

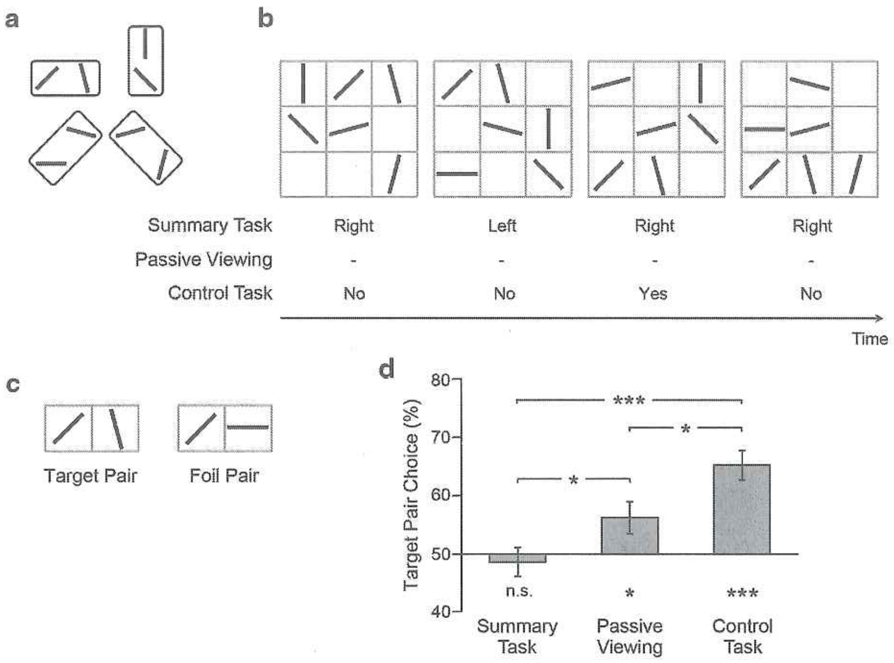

Fig. 5 Statistical summary perception. While being exposed to spatial line pairs, different groups of subjects performed a statistical summary task (*Is the mean line orientation to the left or right of the vertical meridian?*), dual-task control (*Are there any duplicate lines?*), or passive viewing. Statistical learning was blocked by the summary task. (Based on Zhao et al. 2011)

to statistics (e.g., transitional probability matrix, or mean). Thus, statistical learning and statistical summary perception may interact in meaningful ways. We tested what happens to statistical learning when objects are attended during a statistical summary task.

The design of this study most closely matches a spatial visual statistical learning task. Subjects were presented with grids of lines (rather than shapes) of different orientations (Fig. 5). There were eight possible orientations, and unbeknownst to subjects, the orientations were arranged into four spatial pairs. On each grid, three of the four pairs were selected and placed such that at least one other pair appeared adjacently. Thus, the only cue to pair structure was the co-occurrence of particular orientations.

The critical manipulation concerned which task subjects performed during familiarization. One group of subjects was instructed to passively watch the grids, as has been done in prior studies (Fiser and Aslin 2001). A second group of subjects performed a statistical summary perception task on the grids: they were instructed to compute the mean orientation of the lines, and to discriminate this mean as being to the left or right of the vertical meridian. A third group of subjects performed a control task to ensure that differences between the passive viewing and statistical summary tasks could not be attributed to merely needing to perform a second task (which could perhaps enhance attention in a generic way). The control task group determined whether each grid contained a duplicate line. The displays were identical in all conditions, and thus any consequences of the three tasks must reflect the impact of different task sets on statistical learning.

Statistical learning was robust in the passive viewing and control task conditions, but was weaker—actually, non-existent—in the summary task condition. This suggests that computing summary statistics may interfere with statistical learning, possibly because of a reliance on shared statistical computations, or because of a reliance on different spatial scales of attention. That is, global attention helps summary performance (Chong and Treisman 2005), while local attention may be necessary for stimulus-specific learning. In any event, these results provide a further constraint on when statistical learning takes place, and suggest an unforeseen connection between two varieties of statistical processing that had been studied separately.

Aside: The Quest for a Great Cover Task

The findings presented so far demonstrate that what you are doing determines whether you will learn. This fact shows the importance of an often-overlooked methodological aspect of studies on statistical learning: the task that subjects perform during familiarization. The earliest studies of statistical learning employed no task at all during familiarization (Saffran et al. 1996b; Fiser and Aslin 2001, 2002). Passive listening or viewing is not without merit: as we have already seen, certain tasks may be detrimental to statistical learning (Zhao et al. 2011). At the same time, anybody who has run a statistical learning experiment with passive viewing knows the inherent awkwardness in asking subjects to "simply watch". Subjects seem puzzled by such instructions ("how can I do nothing"), and often suspect that the experimenter has ulterior motives. This may lead subjects to implement idiosyncratic strategies in searching out the meaning behind the displays. While such searches rarely turn up the true structure, they may nevertheless be detrimental to learning. In other words, not giving subjects a task does not mean that they won't impose their own task, and, without careful debriefing, such tasks are unknowable and uncontrolled from the experimenter's perspective.

By analogy, the study of functional connectivity in the brain is typically conducted while subjects rest in a functional magnetic resonance imaging (fMRI) scanner with no task (Fox and Raichle 2007). Because no task has been imposed, it is assumed that any resulting patterns of brain activity reflect the stable intrinsic functional architecture of the brain. Yet, recent tasks seep into resting state brain activity (e.g., Stevens et al. 2010), suggesting that subjects are not performing no task (or a default task), just that the experimenters typically do not know what it is.

Thus, a certain amount of experimental control is gained by asking subjects to perform a task during familiarization. It is worth clarifying that such tasks are never to learn the structure of the display (i.e., intentional learning). Not only would this kind of explicit task violate the unsupervised spirit of statistical learning, intentional learning tasks are not always beneficial to learning: in a classic artificial grammar learning study, intentional learning instructions *hurt* learning (Reber 1976). Instead, tasks during familiarization can be considered 'cover' tasks in that they do not themselves disclose the presence of regularities.

When designing studies of statistical learning for which the nature of the cover task does not matter *per se*, one often simply wants to ensure that statistical learning takes place. From many successful and unsuccessful experiments, it seems that the most conducive tasks to learning are those that emphasize the identity of the objects being perceived. For example, the interleaved one-back task described above requires maintaining each shape in working memory, and judging whether the next shape has the same identity. Other tasks that emphasize object identity work well too as shown below, such as detecting motion jitter (requires some shape processing to resolve motion correspondence) and classifying objects into semantic categories. Tasks that do not require any object recognition are less conducive to learning, such as requiring detection of fixation luminance changes. Whether particular tasks are conducive (boost statistical learning with respect to some baseline) or permissive (prevent the blocking of learning that other intrinsic tasks might cause) remains to be further clarified. However, the big picture of why some tasks are good and others are bad has more than methodological implications: humans may be able to control when statistical learning takes place simply by engaging in certain behaviors, without even realizing that they possess this power.

What Happens During Learning?

Conventional statistical learning designs all have something in common: they require two parts. In particular, statistical learning occurs during the first part (familiarization), and then is tested in the second part (test). A separate test is needed because of the canonical use of passive viewing during familiarization. The test is often a two-alternative forced choice familiarity task in which subjects must discriminate between a regularity from familiarization and a foil constructed from the same elements arranged into a new sequence or configuration. While this kind of test has been used successfully in many studies (including many of our own), this design has two drawbacks: First, supposedly incidental and unconscious statistical learning is being tested with an explicit familiarity judgment. While familiarity can be informed by implicit processes, a familiarity test may not be the most sensitive measure of implicit learning. Second, testing learning after the fact means that information about what happens during learning (e.g., about the timecourse of learning) is lost. Indeed, by using two parts, one makes assumptions about how much exposure is needed for learning since familiarization must end before the test can begin—familiarity cannot be measured at multiple intervals, since this would explicitly cue subjects to the presence of regularities.

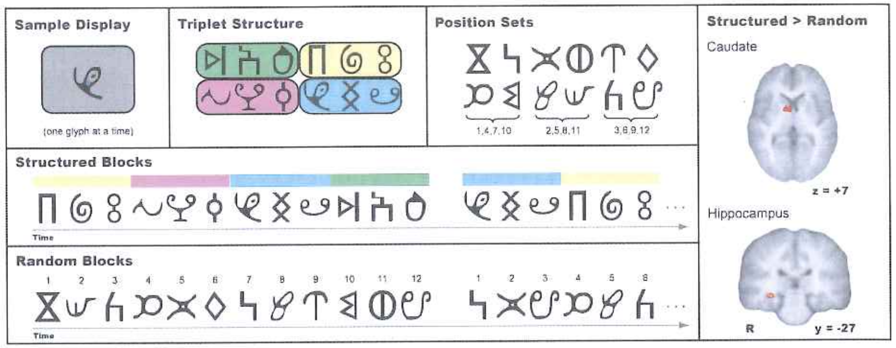

Fig. 6 Neural evidence. Subjects were exposed to blocks of glyphs constructed from triplets or from otherwise matched but random sequences. Both the caudate and posterior hippocampus responded more strongly to structured blocks, providing evidence of statistical learning after only a handful of triplet repetitions. (Based on Turk-Browne et al. 2009)

To resolve these issues, we conducted an fMRI study of what happens in the brain during familiarization (Turk-Browne et al. 2009). We sought to answer three questions about the process of statistical learning in this study: What is the relationship between statistical learning and other forms of learning? How efficient is statistical learning? What is the relationship between our online measure of statistical learning in the brain and subsequent familiarity? Using fMRI, we monitored changes in the brain related to statistical learning without requiring an online behavioral measure (e.g., Hunt and Aslin 2001), and perhaps before these changes manifest in behavior.

Subjects were presented with blocks of nonsense glyphs from ancient alphabets, where each block contained 12 glyphs (Fig. 6). Their task was to detect whenever one of the glyphs jiggled on the screen (a subtle horizontal motion). The blocks alternated between two sets of 12 unique glyphs. One of the sets, which was used to generate the Structured blocks, was composed of four triplets of glyphs. In generating each Structured block, the triplets were randomly sequenced in an order that had never previously been seen. The glyphs appeared sequentially, one at a time, and thus triplets could only be discriminated based on the higher transitional probabilities for glyphs within vs. between triplets. As a comparison, the other set of glyphs, which was used to generate the Random blocks, was composed of three ‘position sets’. Namely, four of the glyphs could appear in the same serial position as the first item in each triplet, four other glyphs could appear in the same serial position as the second item in each triplet, etc. Other than these constraints, the order of glyphs was randomized in each Random block. The overall block sequence alternated between Structured and Random blocks, and we contrasted neural responses to the two block types as a measure of learning. The Structured and Random blocks were identical in terms of item frequency and serial position frequency, and thus any neural differences must reflect sensitivity to the stronger transitional probabilities within triplets in the Structured blocks.

We were especially interested in whether statistical learning would engage well-known memory systems involved in other types of learning. Two systems of particular interest included the medial temporal lobe and the striatum. The medial temporal lobe, and the hippocampus in particular, is the primary brain system for declarative memory (Cohen and Eichenbaum 1993), and has been implicated in some implicit forms of relational (Ryan et al. 2000), configural (Chun and Phelps 1999), and sequence learning (Schendan et al. 2003), which may all be related to statistical learning. The striatum, and the caudate in particular, is often linked to non-declarative or procedural learning, and has been implicated in many forms of implicit learning, including category learning (Seger and Cincotta 2005), artificial grammar learning (Lieberman et al. 2004), and motor sequence learning (Toni et al. 1998). After a few minutes of exposure, the hippocampus and caudate responded more strongly to the Structured blocks than to the Random blocks. These findings suggest potential connections between statistical learning and other forms of learning and memory. We are actively exploring the role of the MTL in statistical learning, for example, examining whether representations in MTL cortex are tuned based on regularities (see Miyashita 1993).

To examine the timecourse of learning, we explored when the difference between Structured and Random blocks emerged during learning using smaller windows of time. In the caudate and hippocampus, as well as other regions, we found initial evidence that statistical learning can occur very quickly, after only 2–3 triplet repetitions. To examine the relationship between this neural evidence of statistical learning and the conventional familiarity measure used in previous studies, we repeated our analyses including only those subjects who performed at or below chance on the familiarity test. Despite not exhibiting any statistical learning by conventional standards, the caudate in these subjects nevertheless discriminated Structured vs. Random blocks. This result suggests that neural evidence of statistical learning can exist without explicit familiarity, perhaps preceding behavioral expressions of learning. In sum, this study uncovered some of the dynamics of learning and identified the neural systems involved.

How Does Learning Handle Real-World Complexity?

We have so far considered when statistical learning takes place and what happens during statistical learning. In this and the next section, I describe what is represented in memory as a result of statistical learning. This question becomes salient when we consider the types of real-world objects over which statistical learning operates. In contrast to the monochromatic shapes used in the prior studies, real-world objects are complex, or ‘multidimensional’. For example, an object can have many features, including color, texture, and shape, and these features can vary over the different parts of an object. How does statistical learning scale up to more complex stimuli? More specifically, when confronted with regularities of complex objects such as colorful shapes, what is learned: patterns of bound objects (e.g., sequences or configurations

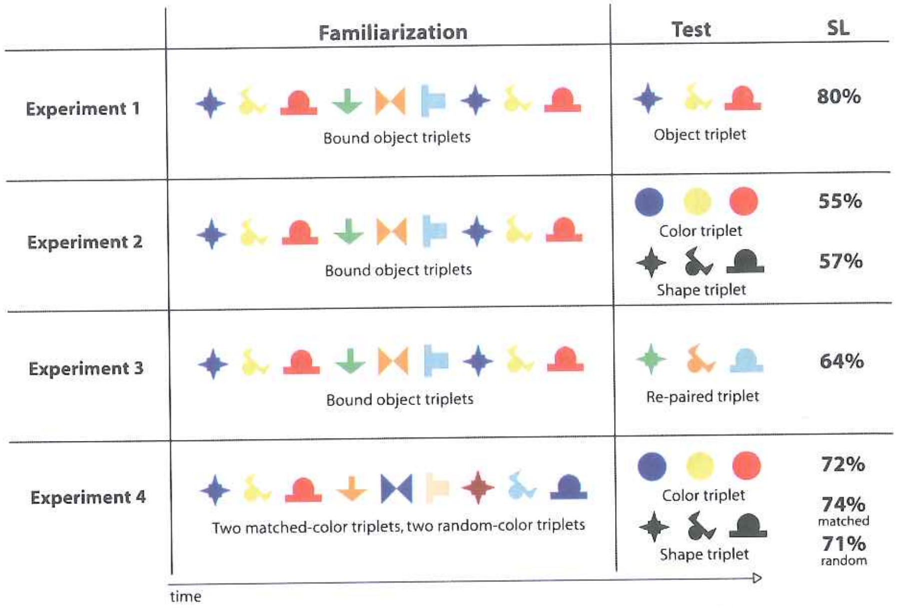

Fig. 7 Multidimensional objects. When exposed to a stream of shape-color objects (each shape has a unique color), statistical learning is better expressed to the bound object (Experiment 1) than to separated features (Experiment 2), or a recombination of features (Experiment 3). When mapping between shape and color is variable, learning is robust for features (Experiment 4). (Based on Turk-Browne et al. 2008)

of shape-color conjunctions) or patterns of the constituent features (e.g., sequences or configurations of shapes or colors)? The answer is not obvious. On one hand, objects provide a strong organizing system for features, and object-based effects have been observed in many domains (e.g., Luck and Vogel 1997; Scholl 2001). On the other hand, some features are not stable over time (e.g., luminance, shadows) and thus learning over conjunctions may be inefficient.

We examined these issues in a series of behavioral studies that looked at statistical learning for multidimensional objects (Turk-Browne et al. 2005). In a temporal visual statistical learning task, subjects were shown a continuous stream of shape-color objects (Fig. 7). Each of twelve shapes was assigned a unique color, and these shape-color conjunctions were grouped into four triplets without subjects' knowledge. To examine whether statistical learning operates over bound objects or separated features, we used a transfer logic: if triplets of bound objects are learned, then familiarity for object triplets should be high at test and familiarity for triplets of the separated shape or color features should be low. Instead, if triplets of separated features are learned, then familiarity for the feature triplets should be robust. The result was that statistical learning extracted the regularities between objects: familiarity was much higher for triplets of bound objects than for triplets of separated features, or even triplets created at test from new combinations of feature triplets.

We further hypothesized that non-diagnostic features should be discounted in learning about objects. For example, certain objects have canonical colors (e.g., fruits and vegetables) while others do not (e.g., furniture, books, hats). We examined whether feature triplets would be better learned when combined into an object with variable or non-diagnostic features from another dimension. We assigned two shape triplets to have fixed and unique colors (as before), and the remaining two shape triplets to appear in randomized colors. Statistical learning now occurred for the shape triplets that were paired with variable colors. Surprisingly, the shape triplets paired with fixed colors were also now better learned, as were the color triplets that appeared with these shape triplets. These findings suggest that the general covariance between feature dimensions determines whether statistical learning outputs knowledge about object or feature regularities: when covariance is high, regularities are learned at the level of objects; when covariance is low, regularities are learned at the level of features.

Another way to think about these results is that statistical learning may always be object-based, but that feature diagnosticity determines what counts as an object: when colors are diagnostic of shapes (and vice versa), objects consist of both a shape and color; when colors are not diagnostic of shapes, objects consist of a single shape or color feature. This interpretation suggests that statistical learning can operate within objects—in addition to between objects—extracting relationships between features (and feature dimensions) to determine which conjunctions are reliable. An object-based bias, coupled with a sensitivity to feature covariance, may allow statistical learning to operate in natural contexts containing regularities among complex objects.

How Flexible Is Learning?

The research above raises a broader question: after learning regularities, how flexible are the acquired representations to changes in the appearance of objects and/or their relationships? In other words, what kinds of changes in the environment can be tolerated when expressing knowledge about regularities? One type of change that has already been discussed is color: we found that statistical learning of colored shapes can be expressed despite the removal of color, but only when colors are non-diagnostic. Another common type of change relates not to the features of individual objects, but rather to the spatiotemporal patterns in which multiple objects appear. Indeed, the visual environment is highly dynamic, meaning that sequences and configurations change over time. For example, the set of people we encounter at the office is relatively stable, but the particular sequence and locations in which we encounter them may vary from day to day. Moreover, because a major source of temporal information in vision comes from eye movements, changes in how we fixate in a given environment (e.g., entering through the back door of a house instead of the front door) will lead to very different sequences. If statistical learning is highly specific to the spatiotemporal details of experience (e.g., Jiang and Song 2005), such variability would be disastrous to the expression of learning. Indeed, given the inherent variability of the world, statistical learning may not be of much use if it cannot cope with some changes between when regularities are acquired and expressed.

We examined these issues in a series of behavioral studies (Turk-Browne and Scholl 2009) where changes were made between the familiarization and test phases of visual statistical learning tasks. Specifically, we examined the flexibility of statistical learning (i.e., generalization) using a similar 'transfer' logic to above: if statistical learning can be expressed despite some change at test, then the changed aspect of the displays is not an integral part of the learned knowledge.

We first examined whether statistical learning could be expressed at test for triplets that appeared in a new temporal order. Subjects watched a continuous stream containing four triplets in the familiarization phase. We then tested whether temporal order was an important part of what they had learned by including both the original (forwards) triplets and, on other trials, reversed (backwards) versions of the same triplets. Insofar as temporal visual statistical learning can generalize across temporal order, subjects should exhibit some familiarity with the backwards triplets. Surprisingly, they expressed equally strong familiarity with the forwards and backwards triplets. This suggests that the knowledge generated by statistical learning is invariant to order.

Since forwards and backwards triplets elicited the same amount of familiarity, were they even distinguishable? Was all order information lost? To test this, we ran an additional study in which subjects were asked to discriminate forwards vs. backwards triplets directly (rather than forwards and backwards triplets vs. nonword foils). Subjects now expressed greater familiarity with forward triplets. These results suggest that statistical learning generalizes across temporal order when order is not needed to discriminate triplets vs. foils, but that this information is represented nonetheless and can be accessed when necessary (when two alternatives can only be discriminated based on order).

Since temporal order information was easily discarded during statistical learning, we then tested whether temporal information was necessary at all for learning to be expressed at test. Following the standard temporal statistical learning task, subjects were tested on triplets vs. foils, but now the objects in the triplets were presented simultaneously in a spatial configuration (Fig. 8). If statistical learning can generalize over the time dimension entirely, subjects should be able to discriminate spatial triplets vs. foils. This is what we found, even when the test displays were presented too quickly for eye movements. These results suggest that temporal visual statistical learning may be useful in building up spatial representations.

Finally, we also examined the flexibility of spatial visual statistical learning. We asked whether learning of spatial pairs would transfer to temporal displays, just like temporal triplets transferred to spatial displays. Saving the details for a little later, the basic result was that spatial learning led to temporal processing benefits. These findings suggest that statistical learning produces knowledge that can be applied flexibly in new contexts. In sum, the studies reported in this section show the power of statistical learning, and its fit to the constraints and properties of the natural environment.

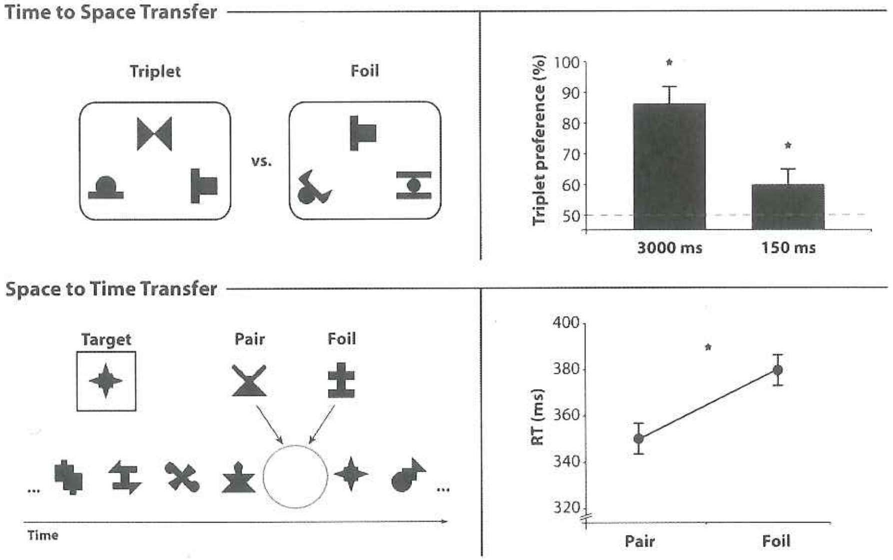

Fig. 8 Spatiotemporal transfer. After temporal visual statistical learning, triplets can be recognized in spatial configurations, even when they are presented too briefly for eye movements. Moreover, after spatial visual statistical learning, pairs lead to RT priming effects in time. (Based on Turk-Browne and Scholl 2009)

Consequences of Statistical Learning

What is statistical learning good for? Many of the studies discussed thus far have relied on familiarity as a measure of statistical learning. Does statistical learning only result in increased familiarity with regularities? This seems somewhat epiphenomenal, in the sense that fleeting familiarity signals may not have much functional significance for ongoing behavior. In this section, I present three other possible consequences of statistical learning, and consider how they might relate to visual search. While there has not been much research specifically testing how statistical learning and visual search interact, the results presented here suggest that such an investigation could be fruitful.

Anticipation

One potential consequence of statistical learning that is often emphasized in the application of statistical learning to language is 'segmentation'—the use of regularities to parse the world into meaningful lexical units or chunks (Giroux and Rey 2009). Similar approaches have been used to model visual statistical learning (Orbán et al.

2008). However, both familiarity and segmentation are *retrospective*, requiring entire patterns to be experienced before they can occur. That is, the relevant kind of familiarity and the ability to place a boundary occur only after an entire regularity has been perceived. Indeed, a hallmark of chunking models is that chunks are indivisible into their component parts (Orbán et al. 2008). Such rigidity may help us recognize higher-order chunks, such as scenes, events, and words, but is not well suited to the dynamic nature of perception. We have thus become interested in whether statistical learning can be used *prospectively*.

In a recent fMRI study (Turk-Browne et al. 2010), we examined what happens at the beginning of temporal regularities—can the brain use knowledge of regularities to anticipate upcoming, predictable objects? For example, imagine repeatedly meeting a particular host when entering a restaurant before being led into the dining room; over time, does perceiving the host's face trigger automatic anticipation of the layout and content of the dining room? In other words, can we rely on learned regularities to live with one foot in the perceptual future? Such anticipation could have important consequences for perception, allowing us to more quickly recognize objects and cope with occluded or degraded visual input.

Subjects in our study were presented with a continuous stream of face and scene images, appearing one at a time and separated by a jittered inter-trial interval (for fMRI analysis purposes). Subjects' task was to decide for each image whether it depicted a face or a scene. Unbeknownst to subjects, eight of the 12 images in each scanning run were divided into four cross-category pairs: two face → scene pairs and two scene → face pairs. The remaining four images (two faces, two scenes) were unpaired, and served as a baseline. We were interested in three conditions: the First image in each of the pairs, the Second image in each of the pairs, and the Unpaired images. Insofar as statistical learning affords anticipation, we predicted that: (a) First images would elicit anticipatory responses compared to Unpaired images, and (b) this anticipation would lead to facilitated processing of Second images.

This latter prediction was apparent in RTs: subjects were faster to categorize the Second vs. Unpaired images, suggesting that the First image had caused priming. Such associative priming effects, resulting from statistical learning, have been observed in two of our other studies as well (Turk-Browne et al. 2005; Turk-Browne and Scholl 2009). In both of these cases, the task on each trial required detecting a pre-specified target embedded in a rapid serial visual presentation stream. We manipulated where the target object appeared with respect to preceding items. For example, in studying whether spatial visual statistical learning transfers to the temporal dimension, the target object was sometimes preceded by the object that it was paired with in space during familiarization and sometimes by an equally familiar foil object. RTs were faster when the target was preceded by its spatial pair, providing evidence that spatial learning can produce temporal cueing benefits. Such effects of statistical learning on object detection and discrimination highlight an important behavioral consequence of statistical learning for object recognition.

Back to the fMRI study of anticipation. Interestingly, RTs to the First images were slower than to Unpaired images. We interpreted this as evidence of anticipation: when perceiving an object that affords predictions about what will come next, this

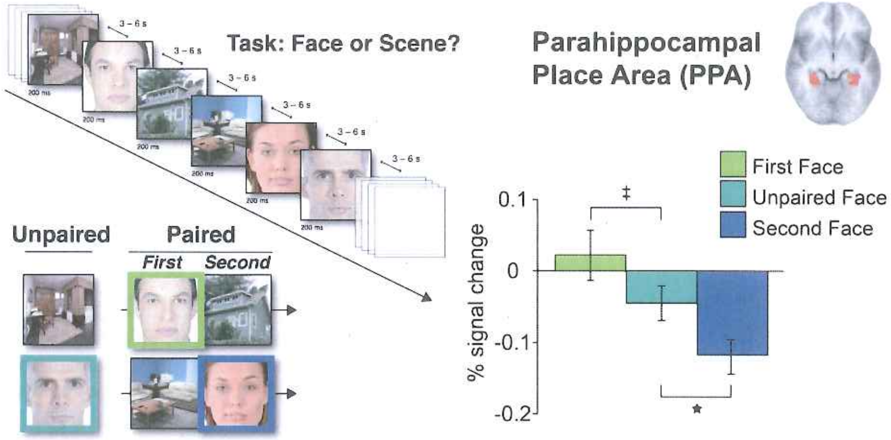

Fig. 9 Implicit perceptual anticipation. Objects that are predictive of what will appear next in the trial sequence produced robust hippocampal activity (not shown). Faces, which otherwise do not activate the scene-selective PPA (see Unpaired condition), elicited enhanced PPA responses when they predicted a scene, and suppressed PPA responses when they could be predicted as non-scenes. (Based on Turk-Browne et al. 2010)

predictive response may act as an implicit dual task and interfere with the current task of categorizing the image in front of you. We are following up on this effect, to determine whether it reflects a generic cost of anticipation (i.e., that performance is slowed whenever anticipation is possible), or prospective response conflict (i.e., that the response for the anticipated image interfered with the required response for the current image). The current study was equivocal: because pairs always contained objects of different categories (for reasons to be described shortly), the response to the First image always conflicted with the response to the anticipated Second image.

Along with these behavioral results, we observed a robust and selective fMRI response in the right hippocampus to the First images compared to the Unpaired images. It is worth emphasizing that First and Unpaired images are identical in all respects but one: they are presented an equal number of times and neither image type is itself predictable based on which image came before; but critically, First images are predictive and allow anticipation of the Second images, while Unpaired images afford no such predictions. Surprisingly, the hippocampal response was evident after fewer than six repetitions of each pair, again providing evidence for the remarkable speed of statistical learning. Finally, although the hippocampus has been traditionally linked to declarative or explicit forms of memory (Cohen and Eichenbaum 1993), subjects reported no awareness that pairs even existed in the stream, suggesting that these anticipatory effects reflect *implicit* perceptual anticipation.

We were especially interested in whether implicit perceptual anticipation could influence visual cortex; in particular, whether regularities could be exploited to potentiate visual processing (Fig. 9). To examine this possibility, we localized category-selective parts of ventral temporal cortex that responded selectively to faces and scenes. If regularities can modulate the visual system, then faces that are predictive of scenes may prospectively elicit scene-related activity (and vice versa). Indeed, scene-selective cortex (the parahippocampal place area, PPA; Aguirre et al. 1998; Epstein and Kanwisher 1998) provides a clean test of this hypothesis since it does not otherwise respond to faces. This can be seen in the response of the PPA to Unpaired faces, which was slightly below baseline. Critically, First faces, which reliably predicted that a particular scene would appear next, elicited an enhanced PPA response. Along with this enhancement, the PPA response to Second faces, which were predictable as non-scenes, was suppressed. These findings provide evidence that anticipation based on statistical regularities can prospectively alter visual cortex.

The idea that statistical learning can prime the detection and recognition of predictable or probable objects may have important consequences for visual search. Indeed, search is speeded when the configuration or identities of distractors predict the identity of the target during visual search (Chun and Jiang 1999; Endo and Takeda 2004). Such effects may result from an effect of regularities on biased competition (Desimone and Duncan 1995), whereby perceiving one object may prioritize processing for associated objects. A related alternative is that regularities could transiently increase familiarity for associated objects, with increased target familiarity leading to more efficient search (Flowers and Lohr 1985).

Relatedly, semantic regularities influence visual search by guiding eye movements to locations in scenes where the target is semantically licensed or probable (Henderson et al. 1999; Neider and Zelinsky 2006; Torralba et al. 2006). For example, receiving a scene preview without a search target facilitates subsequent search when the target is introduced, ostensibly because the scene context in the preview could be leveraged to restrict search to likely target locations (Hollingworth 2009). The brain seems keen on generating predictions, and if those predictions are grounded in regularities, searching for a target object may be aided by perceiving and exploiting associated distractors.

Shifting Modes of Attention

The efficiency of visual search depends on the extra time needed to find a target item for each additional distractor added to a search set, quantified as the slope of RT as a function of set size. Efficiency varies a lot depending on the nature of the target and distractors (Wolfe 2001). Search is most efficient (shallow slopes) when the target differs from all distractors in terms of the presence of one simple feature (Treisman and Gelade 1980). Such targets may be detected pre-attentively, reflecting parallel processing of all items in the display—a 'distributed mode' of attention (Chong and Treisman 2005). In contrast, search is least efficient (steep slopes) when the target and distractors share features, and the target is defined by the conjunction of features (Duncan and Humphreys 1989). Such targets require serial processing of each item in the display—a 'focused mode' of attention (Treisman and Souther 1985). Search efficiency is thus better when a target can be found with distributed attention, and worse when focused attention is required.

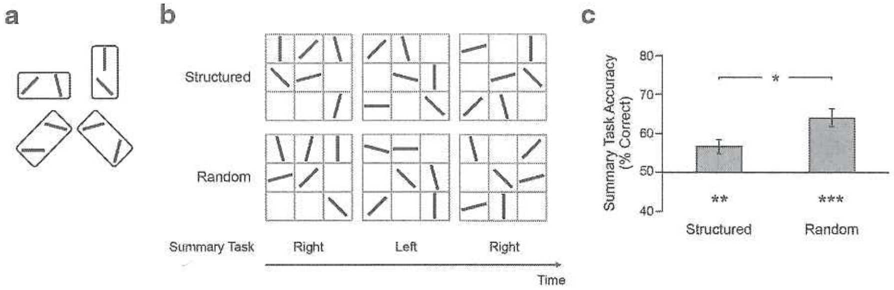

Fig. 10 Regularities and summary perception. Subjects performed a summary task (judging the mean line orientation) over grids that contained spatial regularities (Structured) vs. grids that were otherwise matched but did not contain spatial regularities (Random). Summary perception improved when regularities were removed. (Based on Zhao et al. 2011)

The need for distributed vs. focused attention during visual search is typically attributed to properties of targets and distractors. However, statistical regularities themselves may also control whether attention can be allocated in a distributed or focused manner. Specifically, we are exploring the hypothesis that regularities among local objects draw attention into a more focused mode. While current evidence for this hypothesis is sparse, the idea grew out of a recent behavioral study (Zhao et al. 2011). In that study, we examined how statistical summary perception, a process that depends on distributed attention (Chong and Treisman 2005), is affected by statistical learning. This is the counterpart to an experiment reported earlier on the reverse—how statistical learning is affected by statistical summary perception. The study is described below before I return to the question of how regularities affect attention.

Subjects were presented with grids of lines, and judged whether the mean line orientation was to the left or right of the vertical meridian (Fig. 10). One group of subjects (Structured) received grids that contained spatial pairs. Another group of subjects (Random) received the same grids, but where the locations of lines were shuffled on each trial to destroy the spatial regularities. We predicted that the mere presence of regularities might prompt the visual system to attempt to learn, which may in turn interfere with summary perception. This is exactly what we found: Judgments of mean line orientation were less accurate in the Structured condition.

We conducted a follow-up study to verify that statistical learning *per se* was interfering with summary perception, and not that it was more difficult to summarize the Structured vs. Random displays because they differed in some unintended way. In this new study, one group of subjects was pre-exposed to Structured displays during the duplicate detection control task described earlier (allowing for pre-learning of the line pairs), while another group performed the control task over Random displays (resulting in the same item familiarity, but no knowledge of the line pairs). Both groups then performed the summary task over Structured displays. The group that had pre-learned the pairs gave more accurate summary judgments than the group that was being exposed to the pairs for the first time, presumably because the latter group was engaged in statistical learning while the former group was not. Critically, the displays were identical during this phase, eliminating the possibility of stimulus differences, and supporting the claim that statistical learning interferes with summary perception.

Given that distributed attention is necessary for summary perception (Chong and Treisman 2005), one interpretation of our results is that regularities (even if not learned successfully) shift attention to a focused mode. There are other potential interpretations too, such as that statistical learning and summary perception rely on shared statistical computations, and that engaging in one process interferes with the other by blocking necessary resources. It will therefore be important for future experiments to test the idea that local regularities attract focused attention. Nevertheless, these findings demonstrate that statistical learning can have *costs* for other cognitive processes, along with the benefits for familiarity, segmentation, and anticipation described earlier.

Biasing Locations in Space

In typical visual search tasks, the target location on any given trial is random. When target locations are not random, search performance improves. For example, when target location is predictable from past experience with a specific configuration of distractors, the target can be found more quickly (e.g., Chun and Jiang 1998). Moreover, independently of distractor locations, the probability of targets appearing in specific locations also influences search. For example, when targets appear in one general location on 75 % of trials, target discrimination is facilitated at that location and inhibited at others (Geng and Behrmann 2005; see also Umemoto et al. 2010).

All of these findings suggest that regularities in the locations of targets and distractors can improve search by biasing the allocation of spatial attention. Besides regularities related to the search task, statistical learning may also independently bias the allocation of spatial attention in a way that could influence search. For example, the presence of statistical regularities at one location may draw spatial attention. Consider the natural environment: at any given moment, we are confronted with too many potential sources of regularities and must decide what to learn. We have previously seen that top-down or goal-directed attention to a subset of visual input limits statistical learning to that input. Without such task goals, however, how does the visual system decide what to learn? One possibility is that, like many basic visual cues (e.g., abrupt onsets), regularities themselves act as a cue for spatial attention. Indeed, we have obtained preliminary support for this possibility: targets are detected more quickly when they appear at a spatial location containing temporal regularities, even when the identity, timing, and location of the target are orthogonal to the regularities (Zhao et al., in press).

Another way that statistical learning could influence attention is after learning has occurred: knowing the structure at one location may allow us to divert limited attentional and working memory resources elsewhere in space. We have tested this possibility in a recent study (Al-Aidroos and Turk-Browne, in prep). Subjects first completed a temporal visual statistical learning task in which shapes were presented one at a time at central fixation and subjects detected one-back repetitions. In one condition, the stream was constructed from triplets; in the other condition, the stream was randomized. After some initial exposure to the stream, subjects were instructed to continue detecting repetitions, but also to determine the orientation of rare low-contrast Gabor probes that appeared in the periphery. Probe discrimination was more accurate when the probes were presented during the triplet vs. random streams. Having acquired the triplets during the initial exposure, subjects may have been released from the burden of statistical learning at fixation and better able to monitor the peripheral locations where probes appeared. Statistical learning can thus have diverse effects on spatial attention, and as a consequence, affect other processes that are modulated by attention.

Conclusions

The purpose of this chapter was to describe where regularities exist in the visual environment, to consider ways of studying how regularities are learned, to review recent studies about when statistical learning occurs, what happens during learning, and what is represented as a result, and finally to consider some potential consequences of statistical learning. All of this research highlights the reciprocal connection between perception and memory: statistical learning is an important mechanism for recording visual experience into memory; in turn, learned regularities influence ongoing perception, whether it be object recognition, statistical summary perception, or spatial attention. Given our robust ability to extract and use regularities from the visual environment, statistical learning may play a broad and fundamental role in many cognitive processes, including visual search.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21,* 373–383.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19,* 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition, 53,* 239–261.

Baillargeon, R. (1987). Object permanence in 3 ½ - and 4 ½ -month-old infants. *Developmental Psychology, 23,* 655–664.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science, 3,* 2–13.

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

Beck, M. R., Angelone, B. L., & Levin, D.T. (2004). Knowledge about the probability of change affects change detection performance. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 778–791.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404.

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67,* 1–13.

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

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

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

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System.* Cambridge: MIT Press.

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 B31,* 24–39.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15,* 559–564.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Endo, N., & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception & Psychophysics, 66,* 293–302.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Graf Estes, K. M., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? *Psychological Science, 18,* 254–260.

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

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

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences USA, 99,* 15822–15826.

Flowers, J. H., & Lohr, D. J. (1985). How does familiarity affect visual search for letter strings? *Perception & Psychophysics, 37,* 557–567.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience, 8,* 700–711.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116,* 439–453.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33,* 1087–1116.

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics, 67,* 1252–1268.

Giroux, I., & Rey, A. (2009). Lexical and sub-lexical units in speech perception. *Cognitive Science, 33,* 260–272.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Henderson, J. M., & Hollingworth, A. (2003). Global transsaccadic change blindness during scene perception. *Psychological Science, 14,* 493–497.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210–228.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17,* 273–291.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130,* 658–680.

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

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.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 318–320.

Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance, 33,* 816–828.

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.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–280.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 350–364.

Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience, 16,* 245–263.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46,* 614–621.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607–609.

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

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 USA, 105,* 2745–2750.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 88–94.

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. (1996a). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory & Language, 35,* 606–621.

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,* 27–52.

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience, 11,* 53–60.

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.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25,* 2941–2951.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24,* 1193–1216.

Stevens, W. D., Buckner, R. L., & Schacter, D. L. (2010). Correlated low-frequency BOLD fluctuations in the resting human brain are modulated by recent experience in category-preferential visual regions. *Cerebral Cortex, 20,* 1997–2006.

Thiessen, E. D., & Saffran, J. R. (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.

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.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67,* 867–875.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114,* 285–310.

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 195–202.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.

Turk-Browne, N. B., 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., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21,* 1934–1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience, 30,* 11177–11187.

Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1419–1429.

Watanabe, T., Náez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413,* 844–848.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63,* 381–389.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20,* 423–428.

Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science, 22,* 1212–1219.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (in press). Attention is spontaneously biased towards regularities. *Psychological Science*.