

OXFORD SERIES IN VISUAL COGNITION

Series Editors

Gillian Rhodes

Mary A. Peterson

Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes

Edited by Mary A. Peterson and Gillian Rhodes

Fitting the Mind to the World: Adaptation and After-Effects in

High-Level Vision

Edited by Colin W. G. Clifford and Gillian Rhodes

Human Body Perception From the Inside Out

Edited by Günther Knoblich, Ian M. Thornton, Marc Grosjean, and

Maggie Shiffrar

Understanding Events: From Perception to Action

Edited by Thomas F. Shipley and Jeffrey M. Zacks

Visual Memory

Edited by Steven J. Luck and Andrew Hollingworth

Visual Memory

Edited by

STEVEN J. LUCK AND
ANDREW HOLLINGWORTH

OXFORD
UNIVERSITY PRESS
2008

Chapter 7

Associative Learning Mechanisms in Vision

Marvin M. Chun and Nicholas B. Turk-Browne
Yale University

Visual objects and events do not occur in isolation, but rather within a rich context of other objects and events. Such context is almost always present in everyday vision, and, in the laboratory, has been shown to facilitate the deployment of eye movements (Loftus & Mackworth, 1978; Mackworth & Morandi, 1967; Torralba, Oliva, Castelhana, & Henderson, 2006), object identification (Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989), and visual search (Chun, 2000). For example, the context of a kitchen enhances recognition of a bread box, but not a drum (Palmer, 1975), and a football player is more accurately recognized on a football field than in church (Fig. 7-1; Davenport & Potter, 2004). How and when does context get associated with objects?

First, it is worth noting that distribution of objects in space and time in the visual environment is not random. Although visual cognition typically focuses on the processing of objects in isolation, the visual environment is rife with information about the relationships between objects. In fact, a visual scene can be defined as “a semantically coherent view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner” (Henderson & Hollingworth, 1999, p. 244). In scenes, objects and events tend to covary with each other, and this statistical structure is invariant over time (Chun, 2000; Fiser & Aslin, 2001, 2002). Thus, if perceptual processes are sensitive to this structure, the complexity of processing can be reduced by improving the predictability of information (Gibson, 1963; Gibson, 1966a, 1966b).

What is lacking in theories of vision is a systematic way to characterize the structure of the visual environment. Is there a grammar for vision that can describe where and when objects should appear relative to each other, just as grammar in language dictates how words can be strung together? Although such a descriptive grammar for vision may not exist, there can be little doubt that rich associations exist between object identities distributed in space and time.



Figure 7-1. Scene context effects. In Davenport and Potter (2004), foreground objects (football player and priest, in these examples) were identified more accurately when presented in congruent background contexts (A), than in incongruent background contexts (B). Reprinted with permission from Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564. Blackwell Publishers.

These associative relations may be the syntax for vision—a tool for describing how objects co-occur—which can be exploited to facilitate visual processing. Indeed, the visual system seems designed to encode the associations between objects, and the capacity for such associative learning is prevalent throughout the brain.

In the first part of this chapter, we will review behavioral evidence of associative learning in vision. First, we will consider which types of associations can be learned, and then we will explore the properties and constraints of the mechanisms involved in such learning. We will also discuss how visual processing can be facilitated by knowledge of associative relationships between objects. The second part of the chapter consists of an analysis of the neural mechanisms that support visual associative learning.

7.1 TYPES OF RELATIONAL (ASSOCIATIVE) MEMORY

The visual environment contains a rich mosaic of associations acquired through repeated experience with objects and their contexts. Contextual information comes in a variety of flavors, such as *where* an object is located in space relative to other objects, *which* objects tend to co-occur, and *when* an object appears in time relative to preceding and subsequent objects. How this information is encoded into memory is the domain of associative learning. Specifically, across multiple experiences, associations are formed between those aspects of context that are predictive of a particular location, object identity, or temporal occurrence. These associations are not passive traces of past experience. Rather, they bias

and facilitate future visual processing, helping to anticipate where and when to look, and what to expect.

In everyday vision, the spatial, object, and temporal dimensions of context are all mutually present. This is perhaps most obvious in the case of visual scanning. During scanning, gaze shifts force the visual input to be distributed across time. Each new fixation provides input from a specific location, conveying information about one or more objects. This is also true in the case of visual search while fixated. Covert attention can prioritize objects or locations (and in either case, the other may come along for free). Moreover, according to spotlight and other capacity-limited models of attention (e.g., Posner, Snyder, & Davidson, 1980), not all locations or objects can be sampled in parallel, such that visual input is distributed through time. Thus, in both of these common forms of visual processing, all types of context are simultaneously present.

Many laboratory studies have explored the spatial, object, and temporal aspects of context in relative isolation. It is especially useful to do so when trying to understand the underlying neural mechanisms, since different neural subsystems are specialized for certain types of information. Most of the studies discussed in the next section employ novel or otherwise artificial stimuli to afford greater experimental control and to reduce the influence of prior knowledge on the formation of new associations. Importantly, the insights gained from these artificial displays seem to scale-up to more naturalistic stimuli and situations.

7.1.1 Spatial Associations

Objects and object parts do not appear in random locations, but rather in systematic locations relative to the locations of other parts and objects. As Biederman famously pointed out, one does not experience sofas floating in the sky or fire hydrants on top of mailboxes. Accordingly, observers have greater difficulty recognizing objects that are presented out-of-place (Biederman, Mezzanotte, & Rabinowitz, 1982). However, these context effects may reflect a bias in reporting incongruent information rather than a problem in perception per se (Hollingworth & Henderson, 1998). As a different example, object parts such as the eyes, nose, and mouth always occur in a well-determined spatial relation to each other. Disruption of this configuration, by inverting the face for example, severely impairs face processing and recognition (e.g., Maurer, Grand, & Mondloch, 2002). Spatial associations are even more critical for recognition when an object or part is ambiguous on its own (e.g., Bar & Ullman, 1995). For example, a blurred-out face elicits face-specific neural activity only when placed atop a body (Cox, Meyers, & Sinha, 2004). In the sections below, we will first discuss how such spatial associations are learned, and then focus on an important way in which these associations facilitate visual processing.

7.1.1.1 Spatial statistical learning

At any given moment, we are bombarded with input from the entire visual field. How does the mind segment this input into spatial associations? It turns out that we are remarkably sensitive to statistical regularities between the

relative locations of objects. A compelling demonstration of spatial associative learning employed novel shape displays similar to Figure 7–2A (Fiser & Aslin, 2001). Unknown to the observers, the distribution of shapes in the display was not random. Instead, the displays were structured, such that pairs of shapes (base pairs) occurred in fixed spatial relations (Figure 7–2B). Each display consisted of three base pairs (six shapes), and observers viewed series of displays generated from these base pairs. Importantly, the paired shapes were never presented in isolation; rather, they were always bordered by other nonpaired shapes. Thus, to learn the base pairs, observers had to represent the higher covariance between the paired shapes relative to the other adjacent, unpaired shapes. Observers were exposed to several samples of these displays. Afterwards, to test for statistical learning, observers were required to choose the more

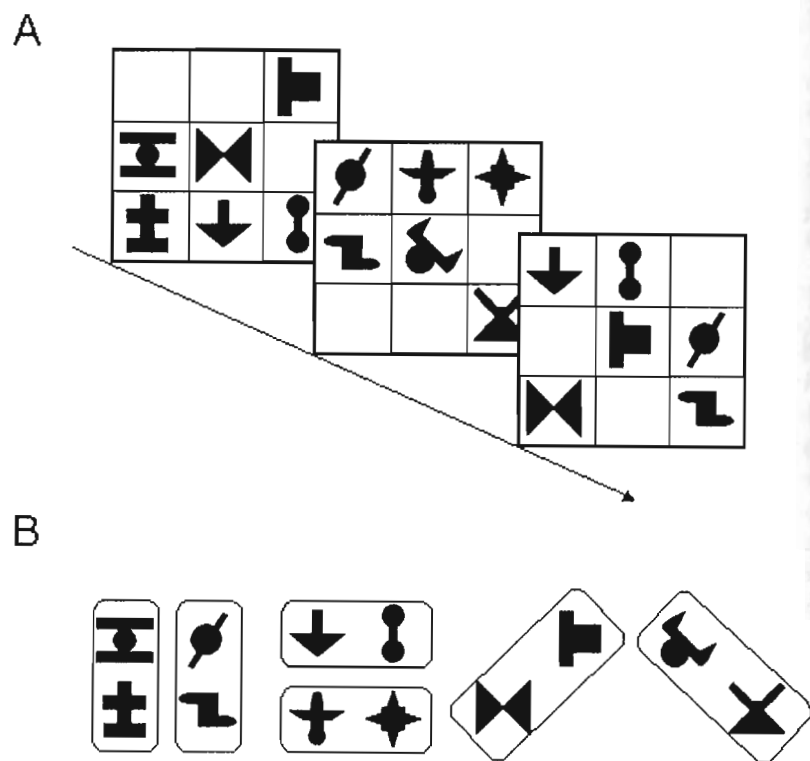


Figure 7–2. Spatial statistical learning. (A) In Fiser and Aslin (2001), observers were presented with grids of shapes during a passive viewing training phase. (B) Each grid was constructed from three “base pairs” — two shapes that always appeared together and in the same spatial layout. However, because the base pairs appeared adjacent to each other, they could only be segmented by relying on higher-order statistics between the shape elements. Adapted with permission from Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6), 499–504. Blackwell Publishers.

familiar of two pairs: One was always a base pair from the first phase, and the other was a new pairing of two old shapes (foil pair). All individual shapes were old, so any familiarity with the base pair must be attributed to spatial association learning. Observers demonstrated remarkably high familiarity with the base pairs.

What information is being extracted in this situation? The most basic statistic that could allow observers to discriminate between the base and foil pairs is joint probability; that is, the relative frequency of co-occurrence. However, in the experiment, the base pairs could also be distinguished based on conditional probability: Given the location of the first shape in a pair, what’s the likelihood that a different shape will appear in a certain adjacent location? To explore which statistical information is being extracted in this situation, Fiser and Aslin (2001) conducted another experiment in which joint probability was controlled, and test items could be distinguished only on the basis of conditional probability. Observers showed sensitivity to the conditional structure of the spatial grids. Further research has suggested that much of statistical learning is the extraction of conditional probabilities (e.g. Fiser & Aslin, 2002, 2005). Such data support the idea that statistical learning mechanisms may provide predictive information in everyday vision.

Most studies of statistical learning have relied on familiarity ratings at the end of the experiment to measure learning. Familiarity tests, however, may not be the best way to assess two important facets of statistical learning. First, to the degree that statistical learning is a form of implicit learning (Perruchet & Pacton, 2006), explicit familiarity judgments may not fully capture the extent of implicit knowledge. Second, testing learning in an offline manner at the end of an experiment could obscure learning-related changes in behavior that might otherwise have been observed online. Motivated by these issues, some studies have employed implicit (e.g., Turk-Browne, Jungé, & Scholl, 2005) and online (e.g., Hunt & Aslin, 2001) measures of statistical learning. Next, we explore a parallel literature on contextual influences in visual search that fully employs online, implicit measures.

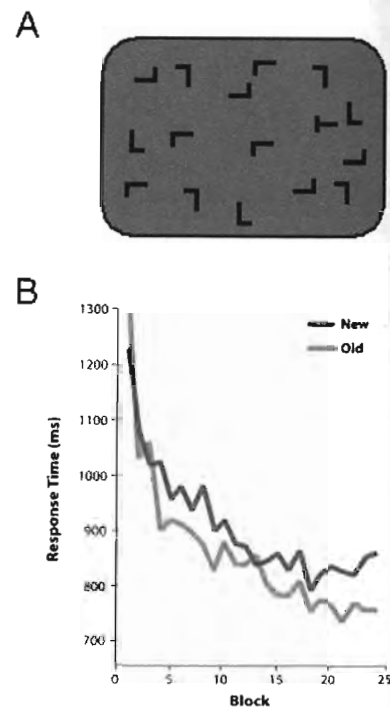
7.1.1.2 Spatial contextual cuing

Studies of spatial statistical learning demonstrate that adult and infant observers have a remarkable ability to segment the visual world into meaningful spatial groups and to associate the items within these groups. But how does such learned knowledge affect visual behavior? The most basic answer is that spatial associations guide where we look. This guidance is necessary to ensure that we prioritize the most salient regions of space. At a traffic intersection, for example, regions of focus should include traffic lights and signs, crosswalks, and the other roads feeding into the intersection. Spatial associations help to filter out stimuli that are intrinsically salient, but in the wrong location. For example, even if a green neon advertisement is flashing in the background, it is critical to focus instead on the red traffic light. Extensive driving experience allows drivers to rapidly orient to and detect information important for driving safely. More generally, spatial context learning is adaptive because major landmarks and the

configurations of various objects in the environment are mostly stable over time, providing useful cues for navigation and orienting. In addition to guiding eye movements, spatial associations can also guide covert attention in visual search tasks. Continuing the automotive theme, consider the instrumentation panel on your car's dashboard. You don't give second thought about where to find the fuel gauge or odometer. The relevant information appears in a fixed location relative to the stable context of the display. Regularities in how information appears relative to the context provide a useful cue for where to attend and look.

To study these intuitions in the laboratory, Chun and Jiang (1998) asked observers to search for a rotated T target amongst rotated L distractors, arrayed in displays as shown in Figure 7-3A. Observers were required to determine whether the T was rotated 90 degrees to the left or 90 degrees to the right. Response time was measured to assess the efficiency of search. Because T's and L's look similar, this type of difficult search task requires careful scanning; The target must be correctly localized with focused spatial attention in order to identify its orientation. The effects of context were tested by simply repeating a set of visual search displays across trials (old condition). These trials were randomly intermixed with trials containing newly generated visual search displays (new condition). In the old condition, the target appeared in a fixed location relative to the configuration of distractors. If observers learned these repeated spatial contexts, attention could be guided to the embedded target location on

Figure 7-3. Spatial contextual cuing. (A) In the standard contextual cuing task (e.g., Chun & Jiang, 1998), observers search for a rotated T target among L distractors, and judge whether the target is rotated to the left or right. Response time is measured as an index of search efficiency. (B) Typical results from a contextual cuing task (from Chun & Jiang, 2003). When the configuration of distractors is repeated across many trials, and consistently paired with a target location (old condition), response times are faster than when the distractor and target locations are randomized (new condition). This difference provides evidence that observers can learn the association between spatial configurations and target locations. Adapted with permission from Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 224–234. American Psychological Association.



subsequent trials. In other words, search should become faster and faster as the old displays are repeated, relative to the new displays in which the context was uninformative. The comparison to new displays is essential because search performance also improves for targets in new displays because of perceptual learning of targets and distractors, and because of practice effects in generating responses to detected targets. This latter form of learning is noncontextual (procedural), while the additional benefit for old displays relative to new displays reflects associative/contextual learning.

Typical results are shown in Figure 7-3B (Chun & Jiang, 2003). Search was faster for targets appearing in old displays than in new displays. This indicates that the context of the target facilitated search for the target by cuing visual attention to the target location. Critically, this cuing effect—known as *contextual cuing*—is a spatial effect because the configuration of distractors did not predict the identity of the target, just its location. In other words, the context predicted where the target could be found, but not what it was. Likewise, contextual cuing is typically not sensitive to distractor identities, encoding just their locations and configurations (Chun & Jiang, 1998; but see Jiang & Song, 2005). We'll consider specific examples of identity cuing in the object association section later in the chapter.

It typically takes about five repetitions of each display (one epoch) for learning of the displays to facilitate search. Interestingly, observers do not usually become aware that some displays repeated or that spatial context is predictive of target location. Such explicit knowledge is formally tested by having observers discriminate old from new displays at the end of the session, and by having them try to guess the target location given an old display in which the target has been replaced by a distractor (Chun & Jiang, 2003). The dissociation between facilitated response times and explicit knowledge suggests that the learning underlying contextual cuing is implicit. Moreover, such knowledge is of high capacity and durable: Observers can implicitly learn dozens and dozens of different layouts while retaining immediate contextual cuing benefits even after a 1-week delay between sessions (Chun & Jiang, 2003; Jiang, Song, & Rigas, 2005).

Given this robust phenomenon of contextual cuing, one may ask whether all information in the background provides equally potent context. In the spatial contextual cuing task, it turns out that there are some constraints on which spatial information can facilitate search. Olson and Chun (2002) discovered that only a portion of each display is associated with a target location. In their study, one side of each display was repeated over multiple trials while the other side changed randomly from trial to trial. Thus, for each old display, half of the display was predictive and the other half was not, and the target could appear on either side. Contextual cuing for targets appearing within the predictive local context was as strong as when the entire display was repeated, suggesting that local context is sufficient; an equally predictive context on the opposite side of the display from the target did not produce contextual cuing. In fact, even a set of only three items around the target within the same screen quadrant was sufficient to elicit contextual cuing (Brady & Chun, 2007; see also Jiang & Wagner, 2004). Together, these results indicate that, in visual search tasks using

dense displays, spatial contextual cuing is locally restricted to the region around the target.

The fact that spatial contextual cuing is driven by local items has several implications. First, this provides useful constraints on learning. Because there is so much information to encode in any background context, the complexity of the associative learning mechanism underlying contextual cuing is reduced by restricting the input. Second, the fact that contextual cuing is driven by local rather than global context sets a limit on how much guidance to expect from predictive spatial contexts in visual search. If predictive global context was implicitly recognized as soon as the display appears, then the associated target should be detected immediately, as would be the case if, for example, the target were a unique, brightly colored, blinking item. However, contextual cueing does not reach such levels of efficiency; the benefit is intermediate between perfect guidance and no guidance (Chun & Jiang, 1998; Kunar, Flusberg, Horowitz, & Wolfe, *in press*). Because predictive context is local, it cannot guide search to an embedded target until the relevant local cluster of items (e.g., quadrant of a display) is reached during scanning of the display. Computer simulations accurately predict how much guidance in search one should observe given the spatially constrained nature of contextual cuing (Brady & Chun, 2007).

Contextual cuing provides evidence of spatial associative learning and shows how such learning can benefit performance. But, how exactly do contextual associations facilitate search? We had proposed that context guides "attention" based on the assumption that the allocation of attention to a target precedes any action directed towards it; that is, faster responses reflect faster allocation of attention to the target location. However, prior studies did not directly demonstrate that the target location is more quickly attended. Moreover, it remains possible that contextual cuing reflects prioritization of the target location at a postselection stage—that is, in visual short-term memory or response selection (Kunar et al., 2007)—rather than prioritized selection of the item at the target location *per se*.

Support for the claim that contextual cuing reflects prioritized sampling of the target location comes from eye movement data, electrophysiological evidence, and functional magnetic resonance imaging (fMRI) studies. For example, fewer saccades are needed to acquire a target appearing in an old display compared to a new display (Peterson & Kramer, 2001a), and the first saccade in a natural scene is most often directed at the probable location of a target (Eckstein, Drescher, & Shimozaki, 2006). Similar results have been observed in monkeys making eye movements to targets embedded in natural-scene backgrounds (Sheinberg & Logothetis, 1998). Interestingly, such contextual cuing of eye movements may even override the powerful pull of salient visual events such as abrupt onsets (Peterson & Kramer, 2001b).

As more direct evidence, event-related potential and fMRI studies have demonstrated that attention is cued by contextual memory. The scalp-recorded N2pc component is a well-validated electrophysiological marker of the focusing of attention (Luck, Girelli, McDermott, & Ford, 1997). If contextual cuing increases the probability that attention is oriented to a target more quickly on

repeated trials, then one should predict an increase in the amplitude of the N2pc waveform. Supporting this prediction, Johnson, Woodman, Braun, & Luck (2007) discovered that the N2pc amplitude was greater for repeated arrays than for novel arrays beginning at a latency of about 175 ms. This finding provides direct evidence that contextual cuing leads to greater early allocation of attention to the target. A separate fMRI study compared visually driven search and memory driven search. In visually driven search, observers were explicitly instructed to orient visual attention to a target location. In memory-driven search, observers learned to orient to a target location based on its associations with a background scene. Frontoparietal regions of the brain important for attentional orienting were commonly activated in both visually driven and memory-driven tasks (Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006). All of these results support a guided attention account of contextual cuing.

Although most of the laboratory studies just discussed employ highly artificial or abstracted displays, contextual cuing also operates in more natural displays. For example, contextual cuing is robust in three-dimensional displays (Kawahara, 2003) or pseudo-naturalistic displays with three-dimensional perspective (Chun & Jiang, 2003). Further, the arbitrary location of a target letter can be associated with the layout of a real-world scene (Brockmole & Henderson, 2006). Such scene-based contextual cuing shows some differences compared to contextual cuing in tasks using arbitrary stimuli. For example, observers became explicitly aware of the scene-target associations, and the degree of learning was sensitive to the semantic cohesiveness of the scene. It may be impossible to eliminate the influence of semantic knowledge in processing real-world scenes (and this is part of the motivation for using artificial displays), but such influences should still be considered contextual (Bar, 2004). For example, targets are identified in real-world scenes more quickly when their location is semantically consistent with the scene (Neider & Zelinsky, 2006).

Interestingly, unlike traditional demonstrations of contextual cuing, target locations get associated with global information in natural scenes (Brockmole, Castelano, & Henderson, 2006): Learning of context transfers across local but not global relational changes. Given that contextual cuing in natural scenes varies as a function of semantic cohesiveness (Brockmole & Henderson, 2006), observers may be learning an association between the target location and a more holistic semantic representation of the scene, even if the identity of the target is semantically unrelated. Thus, local changes within a given semantic context will not affect the expression of learning, whereas global changes to the semantics of a scene might. Such an account is consistent with the fact that the parahippocampal place area (PPA), a region of ventral visual cortex selective for natural scenes, is more sensitive to global layout than to object configurations (Epstein & Kanwisher, 1998a). Moreover, processing in nearby parahippocampal cortex may serve as a semantic frame for object recognition (Bar, 2004) and, more specifically, may be directly involved in binding objects to background scenes (Goh et al., 2004). Thus, contextual cuing in natural scenes may reflect simpler associations between holistic scene representations and target locations,

rather than between configurations of distractor and target locations. Such reliance on global features is a key component of a very promising computational model that can produce contextual guidance of eye movements and attention in real-world scenes (Torralba et al., 2006). This elegant and sophisticated model combines bottom-up saliency, scene context, and top-down mechanisms to predict image regions that are likely to be fixated during natural search tasks.

7.1.1.3 Links between spatial statistical learning and contextual cuing

In many ways, contextual cuing can be considered a special case of spatial statistical learning: Observers extract statistical relationships between the configuration of distractors and the target location. In other ways, the two phenomena are different. One superficial difference is that contextual cuing per se reflects the expression of learned associations; thus, the progression of learning can be easily tracked online, unlike most studies of statistical learning (cf. Hunt & Aslin, 2001). A more profound difference is the conscious accessibility of the learned associations: Most studies of statistical learning have relied on explicit familiarity tests to assess learning (cf. Turk-Browne et al., 2005), whereas, in the case of contextual cuing, such tests fail to elicit evidence of learning (e.g., Chun & Jiang, 1998, 2003). Therefore, to the degree that contextual cuing reflects statistical learning, familiarity may not be the most informative measure of statistical learning.

Another important consideration is the distinction between association and segmentation. The result of statistical learning could simply be a set of associations between elements, where the strength varies as a function of the statistical properties of the input. Performance on the familiarity test would then reflect the relatively stronger associations between the elements that had been grouped versus ungrouped. Another possibility, however, is that statistical learning results in segmented units or chunks: “words” in auditory statistical learning (e.g., Saffran, 2001) and “objects” in visual statistical learning (e.g., Fiser & Aslin, 2005). According to this view, the boundaries between chunks may be determined by statistical computations (or independently; see Perruchet & Pacton, 2006), and test performance may reflect the strength of the chunk’s memory trace per se rather than the strength of the associations between constituent elements. Chunking does not seem to apply well to contextual cuing, however, since there is little information to be segmented in any given display; observers may instead learn associations between the target location and individual distractor locations (Jiang & Wagner, 2004).

7.1.1.4 Global spatial context

Thus far, we have focused on context defined as the distribution of objects in space. However, spatial context can also be defined at a stage prior to object individuation. For example, the semantic category of a scene can be determined from low spatial frequencies (e.g., Schyns & Oliva, 1994), the distribution of colors (Goffaux et al., 2005), and the distribution of orientations (McCotter, Gosselin, Sowden, & Schyns, 2005). Attention can then be guided efficiently to the most salient locations in a scene by combining this global semantic

information with bottom-up local salience (e.g., Torralba et al., 2006). One advantage of this global property approach is that individual objects do not need to be segmented and identified. This does not rule out a contribution of individual object recognition to context effects, however. Both global scene-based and individual object-based forms of spatial context information can be integrated in a single model of visual recognition. Low spatial frequencies and spatial layout information in a scene may help activate a semantic frame and identify a range of possible interpretations of a target object, whereas high spatial frequencies may be contrasted with the possible interpretations and constrained by the semantic frame (Bar, 2004; Chun, 2003). Such approaches suggest that visual processing is most efficient when it takes advantage of statistical structure at multiple spatial scales and levels of complexity.

7.1.2 Object Associations

Although context specifies where things can be found, it can also specify what kinds of things should be expected. Objects that tend to covary in the environment can become associated, independent of their spatial relations. For example, when looking at an office desk, one does not expect to see lawnmowers or rakes, but rather paper and pencils, even if their precise locations change from day to day. In this section, we explore how the covariance between objects (independent of spatial location) can be learned, and how this knowledge can be exploited to facilitate object recognition.

7.1.2.1 Object contextual cuing

The learning and expression of object covariance has been investigated using a modified contextual cuing paradigm (Fig. 7–4). Observers were presented with arrays of novel shapes containing a single target shape that was symmetric around the vertical axis and several distractor shapes that were symmetric around nonvertical axes (Chun & Jiang, 1999). Observers were instructed to search for the vertically symmetric object as quickly and as accurately as possible. By defining the search target according to symmetry, novel shape targets could be specified without labeling or showing the actual shape. As soon as observers detected the target, they pressed a key and response time was recorded. The array of novel shapes was then replaced with an array of probe letters, each appearing in a location previously occupied by a novel shape. Observers entered the probe letter that appeared in the same location as the target shape. This non-speeded probe task ensured that the target was localized correctly.

As in the spatial contextual cuing task described in the previous section, two conditions were present. In the old condition, the target shape was always paired with a fixed set of distractor shapes from block to block. In the new condition, the target shape and distractor sets were uncorrelated: The assignment of distractor shapes to a particular trial was randomized independently of the target shape. In other words, the target and distractor shapes consistently mapped in the old condition but, in the new condition, the mapping between target and distractor shapes was variable (Schneider & Shiffrin, 1977; Shiffrin

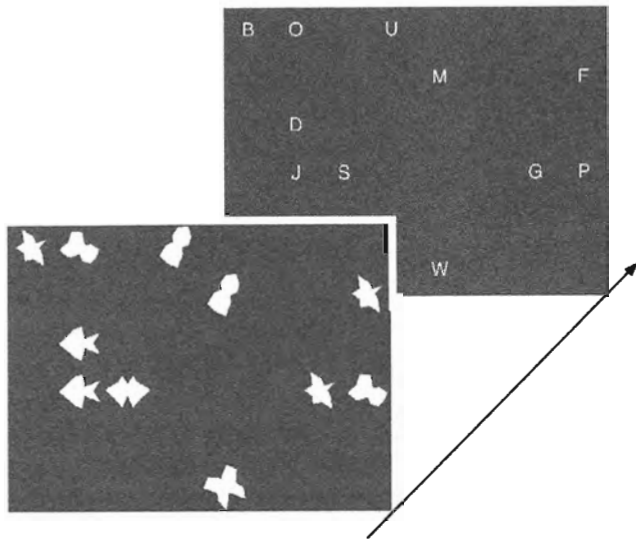


Figure 7-4. Object contextual cuing. In a variant of the contextual cuing task, observers search for the shape with vertical symmetry among non-symmetrical distractors (Chun & Jiang, 1999). To verify that the target shape was correctly localized, observers then respond by pressing the letter in the target location. In the old condition, the symmetrical target shape is consistently paired with a particular set of distractor shapes. In the new condition, the target shape is paired with a random assortment of distractor shapes from trial to trial. Critically, the distractor and target locations are completely randomized for each trial, eliminating the possibility of spatial associative learning. Visual search eventually becomes faster in the old condition than in the new condition. This difference provides evidence that the target and distractor shapes in the old condition have been associated. Reprinted with permission from Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360–365. Blackwell Publishers.

& Schneider, 1977). All target shapes and distractor sets were repeated an equal number of times in both conditions, ruling out perceptual learning of individual shapes as a contributing factor in performance. Critically, the location of the target and distractors in a given display were fully randomized from trial to trial. Thus, any differences in performance between old and new conditions must be attributed to the covariance or consistent mapping between target and distractor sets, divorced from spatial structure (later, in the section on cross-dimensional learning, we'll discuss a study in which both object and spatial statistical structure were simultaneously present).

Observers demonstrated robust object contextual cuing: On trials where the target probe letter was correctly identified, target detection was faster in old displays than new displays. This reveals that we are sensitive to the covariance of objects in the visual environment in spite of (or because of; Endo & Takeda,

2004) spatial variability, and that learned object–object associations can facilitate recognition of search targets in context. Like spatial contextual cuing, such associative learning produces implicit knowledge: Observers performed at chance when attempting to discriminate between old and new displays.

The wealth of research on spatial contextual cuing raises many questions that have yet to be answered about the nature of object contextual cuing. For example, how do the distractor shape identities get associated with the target identities? Specifically, individual shapes may become associated in pairs, and thus, the serial processing of distractors associated with the target shape will increasingly activate the target representation. However, the distractor shapes could also become associated together in a set, which, in turn, would become associated with the target identity; from work on consistent versus variable mapping (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), we know that identity set effects are powerful in visual learning and search.

Finally, it is interesting to consider how the attentional guidance in spatial contextual cuing might relate to the facilitation observed here. Specifically, because the target and distractor locations were randomized, the effect cannot be spatial. How can knowing the identity of the target facilitate search through a series of spatial locations? In serial searches, object contextual cuing may reflect facilitation at either (or both) selection and identification stages. In other words, the target location may be selected more quickly because previously processed distractors would set up an attentional set for the target identity. Additionally, the distractors may not affect the speed with which the target location is selected, but rather, the speed with which the target is identified after selection. According to this view, the distractors may associatively prime the identity of the target, facilitating the decision of whether the selected shape has vertical symmetry.

7.1.2.2 Spatial specificity of learning

Spatial and object contextual cuing tasks were purposefully designed to test for the formation of one specific type of knowledge without influence from the other. In the case of spatial contextual cuing, observers can only learn spatial associations: Only the spatial configuration of distractors is predictive (distractors are homogenous L shapes), and only the target location can be predicted (target identity is chosen randomly). Conversely, in object contextual cuing, observers can only learn object associations: Only the set of distractor identities is predictive (distractor locations are randomized), and only the target identity can be predicted (the target location is randomized). Such restricted tasks were deliberately constructed to isolate specific mechanisms, but may never be fully dissociable in other situations.

For example, both sources of structure are present in spatial statistical learning tasks. In other words, in the study of Fiser and Aslin (2001) discussed earlier, observers could have learned two things about the base shape pairs: (a) that the two shapes are associated in a particular spatial configuration, or (b) that the two shapes are simply associated (with no spatial specificity). Their study was unable to distinguish between the two possibilities because the pairs were

always tested in their original configuration. However, if spatial and object learning can be separated, do associations learned in one dimension generalize to the other?

A recent study demonstrates that spatial statistical learning can result in object associations (Turk-Browne & Scholl, in press). In this study, observers were exposed to the training displays from Fiser and Aslin (2001), but statistical learning was tested in a different way. Observers performed target detection in rapid sequences of shapes presented one-at-a-time in a single location. Detection was facilitated when the target shape was preceded by the shape with which it was spatially associated during learning. Because no spatial structure was present at test, these results suggest that observers were able to abstract object associations in the presence of spatial structure.

Future research could explore the situations under which spatial associations between objects can be abstracted to pure object associations. As demonstrated by object contextual cuing, variability in spatial configurations can encourage learning of object-object associations (see also Endo & Takeda, 2004). Another important factor may be the nature of the task during learning. In contextual cuing, observers are actively engaged in visual search through space; in typical spatial statistical learning tasks, observers passively view displays without a spatial task.

7.1.2.3 Learning within objects

While we have thus far focused on learning of associations between objects, a wealth of information also exists about covariance between object parts (used loosely for both surface and shape features). Much like spatial and object context, feature associations can affect object recognition. For example, it takes longer to name the color of a purple banana than a yellow banana (Naor-Raz, Tarr, & Kersten, 2003), purportedly because the banana shape is associated with the color yellow, which causes Stroop-like interference. The strength of an association between shape and color within an object is governed by *color-diagnosticsity*, that is, the variability of the mapping of a particular color to a particular shape (Tanaka & Presnell, 1999). For example, objects with high diagnosticsity (e.g., fire engines) are more quickly recognized when presented in color, whereas objects with low diagnosticsity (e.g., chairs) are just as quickly recognized without color. These results suggest that colors become associated with shapes because they co-occur in the environment. In turn, such covariation between surface features may help define the objects over which other forms of associative learning operate (Turk-Browne, Isola, Scholl, & Treat, 2008).

In addition to surface features, two or more shape features may themselves become associated into objects on the basis of co-occurrence (Fiser & Aslin, 2005), and grouping cues can further help shape features become associated with each other (Baker, Olson, & Behrmann, 2004). The result of learning within objects may be the attentional prioritization of other regions of the object (Igley, Driver, & Rafal, 1994; Shomstein & Yantis, 2002), analogous to the guidance of attention between objects.

7.1.3 Temporal Associations

In a static snapshot, spatial configurations and object identities provide rich context for visual search and recognition. However, this context can change from moment to moment for several reasons. First, the visual environment itself is highly dynamic; in order to perceive the world with some continuity, we must be able to recognize objects over time and transformations. Second, the visual input received from our eyes is dynamic even when the world is static. We constantly move our eyes and head to sample new regions of space, and such movements cause massive shifts in retinal input. The visual system is remarkably good at compensating for such shifts, helping to maintain a stable percept (this compensation can be subverted by directly moving the eyeball with a finger). Third, because of the capacity limitations of spatial attention and visual short-term memory, we must iteratively sample the visual environment. This serial processing ensures that some spatial and object associations in the environment will be distributed across mental time. Finally, events can unfold over time in predictable ways. This rich temporal structure may guide our expectations for what will happen from one moment to the next. We will consider how observers can learn regularities in how objects move in a later section on spatiotemporal learning. In this section, we will focus on how context distributed in time is learned, and how such context can help predict upcoming events.

7.1.3.1 Temporal statistical learning

As discussed earlier, the distribution of objects in space is often continuous: Objects are associated into groups, but these groups are not necessarily spatially segregated. Rather, we can learn to segment or chunk the visual world into meaningful units via spatial statistical learning. Similarly, visual processing is continuous in time: The focus of attention changes from object to object as we navigate through the visual environment. Embedded within this continuous stream are two types of structure: (a) reliable and ordered sequences of input, or "events" (Zacks & Tversky, 2001) and (b) associations between objects that tend to co-occur in the environment and thus are more likely to be processed in temporal proximity. The ability to extract and represent both kinds of temporal structure is another form of statistical learning and, much like spatial statistical learning, seems to be governed by the conditional probabilities of (now temporally) adjacent items.

To study temporal statistical learning, Fiser and Aslin (2002) presented short animations of simple shapes appearing one at a time (Fig. 7-5A). Unknown to observers, the stream contained very regular patterns. Specifically, the shapes were assigned into groups of three that always appeared in the same order (Fig. 7-5B). The stream was constructed by randomly sequencing these "triplets," such that the transitional probabilities between shapes within a triplet was very high, whereas the transitional probability between shapes at the triplet boundaries was very low. Importantly, the stream was continuous, thus providing no temporal cues to the boundaries between triplets. Learning was

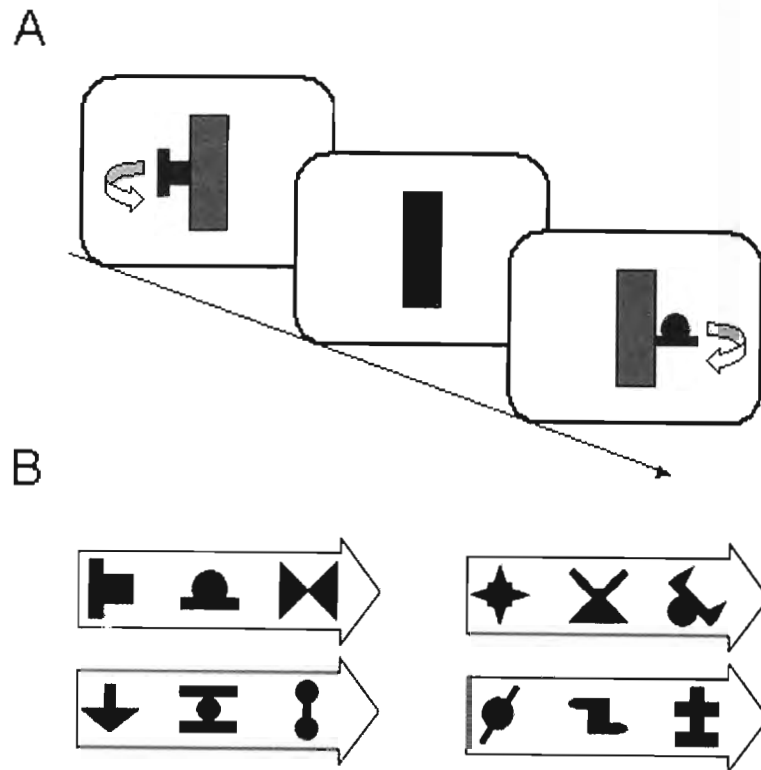


Figure 7-5. Temporal statistical learning. (A) In Fiser and Aslin (2002), observers watched a stream of shapes appearing one at a time from behind an occluder. (B) The stream of shapes was constructed from subsequences of three shapes that always appeared in succession. Observers learned these subsequences, but because the stream was continuous (and thus the boundaries between subsequences were unmarked), they were required to rely on higher-order statistics between the shape elements. Adapted with permission from 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(3), 458–467, Figures 2 and 3. American Psychological Association.

tested by having observers repeatedly judge whether the triplets or new combinations of the same shapes were more familiar. Because only familiar shapes were used, the two alternatives could only be distinguished based on the relatively higher transitional probabilities within the triplets. Observers demonstrated a robust familiarity preference for the old sequences, suggesting that they had learned the triplet structure of the stream. This segmentation may reflect sensitivity not only to the high transitional probabilities between shapes within a triplet, but the high variability of the transitions at the triplet boundaries. Interestingly, high variability alone may also serve to define

boundaries between events in natural vision, even without other cues (Avrahami & Kareev, 1994).

Sensitivity to temporal structure may be domain general. Specifically, one of the most prominent demonstrations of temporal statistical learning was in the domain of language acquisition (Saffran, Aslin, & Newport, 1996). In that study, 8-month-old infants were presented with streams of syllable triplets and, after 2 min, could distinguish between repeated and novel sequences. Such learning may provide an especially useful mechanism for word segmentation, since normal speech is undifferentiated (Saffran, 2001). However, as in the case of spatial statistical learning, observers may not, in fact, be extracting temporal sequences per se (Turk-Browne & Scholl, in press). After being exposed to a stream of triplets (as in Fiser & Aslin, 2002), observers could recognize the triplets in static spatial displays that were presented for only 150 ms. This suggests that, although the processing of objects may be distributed across time, these regularities are represented as simple object associations. Of course, representations may also exist that possess temporal order; such specificity would be critical for language learning, where words are not simply a collection of syllables, but possess a particular ordering.

7.1.3.2 Temporal contextual cuing

These studies of temporal statistical learning rely on familiarity tests after the fact. How does knowledge of temporal sequences affect performance? In another variant of contextual cuing (Olson & Chun, 2001), observers were presented with sequences of letters. The letter identities appeared in either a fixed sequence or a random sequence. The task involved detecting a prespecified target letter as quickly as possible. When the target letter was preceded by a repeated (and hence, predictive) sequence of letter identities, observers were quicker to respond than when it was preceded by a random sequence. This demonstrates sensitivity to the temporal structure of events, such that prior processing can facilitate future recognition. Similar facilitation effects have been observed as a result of temporal statistical learning (Turk-Browne et al., 2005) and appear to grow when additional context is added. Independent of object identity, observers are also sensitive to the timing or rhythm of sequences. Target detection is faster when the target appears at a time that can be predicted by the onset times of the preceding items (Olson & Chun, 2001; see also Shin & Ivry, 2002).

7.1.3.3 Other (more abstract) types of temporal associative learning

Although we are focusing here on visual temporal context, the topic of sequence learning more generally has a long and important history. Much of this research began as an exploration of the distributional properties of language (Harris, 1955). Subsequent research explored the mechanisms of grammar learning, specifically, how experience with a set of exemplars can produce knowledge of the artificial grammar from which they were generated (Reber, 1967). More recently, research has been focused on how sequences of manual responses can be learned (serial reaction time learning; Nissen & Bullemer, 1987). The degree

to which the mechanisms involved in these forms of learning apply to visual associative learning remains an open and interesting question. For example, similar constraints apply to serial reaction time and statistical learning: Selective attention is necessary, but learning can occur without intent or awareness, and even during secondary tasks (Jiménez & Méndez, 1999; Turk-Browne et al., 2005). On the other hand, serial reaction time learning and spatial contextual cuing may rely on different neural mechanisms. For example, contextual cuing is spared in healthy aging, whereas sequence learning can be impaired in the same subjects (Howard, Dennis, Howard, Yankovich, and Vaidya, 2004).

7.1.4 Cross-dimensional Associations

As discussed earlier, the isolation of spatial, object, and temporal aspects of context is theoretically useful, but one must also characterize the wealth of learning that occurs between these dimensions. In everyday vision, these three facets of context are not only simultaneously present, but are often inextricably linked. Thus, we review here some examples of how learning operates over multidimensional or *spatiotemporal* objects.

7.1.4.1 Spatial–temporal learning

Regularities exist in how objects change over time. For example, in many team sports such as basketball, football, and soccer, one must be sensitive to how other players move around the field. These movements are not random but are also not necessarily instructed. Skilled players are able to sense these patterns, which can in turn allow them to anticipate and exploit how a play will evolve. The best players are those with “field sense,” which refers to the ability to make an optimal play based on the global state of how players are moving about. To mimic this situation in the lab, one study asked observers to search for a target that moved around the computer screen amid a set of moving distractors (Chun & Jiang, 1999). All of the objects moved along independent trajectories, with the constraint that they could not run into each other. To study the effects of dynamic context, the target motion trajectories in the old condition were consistently paired with their own sets of distractor motion trajectories. In other words, the entire dynamic vignette was repeated from block to block. In the new condition, the target and distractor motion trajectories were randomized from block to block. Observers were faster to detect the target in the old condition, indicating that they implicitly learned the dynamic contexts that were predictive of the target motion trajectory. This is analogous to how patterns in player movements on a ball field may allow an experienced eye to predict how events will unfold. The associative learning of motion trajectories may be analogous to the learning of object associations; like objects, brief reliable motions are a basic unit of perception (Cavanagh, Labianca, & Thornton, 2001).

Not only can we learn patterns of object motion, we can also learn patterns of where and when different objects will appear during dynamic events. To test this kind of learning in the laboratory, observers can be presented with a series

of objects appearing one at a time in a very regular sequence of spatial locations. In traditional serial reaction time tasks (e.g., Nissen & Bullemer, 1987), observers press a different response key for each of the possible locations, and response times decrease as the sequence is repeated. However, because the spatial sequence and the response sequence are perfectly correlated in this design, it is unclear whether the learning is spatial rather than response-driven.

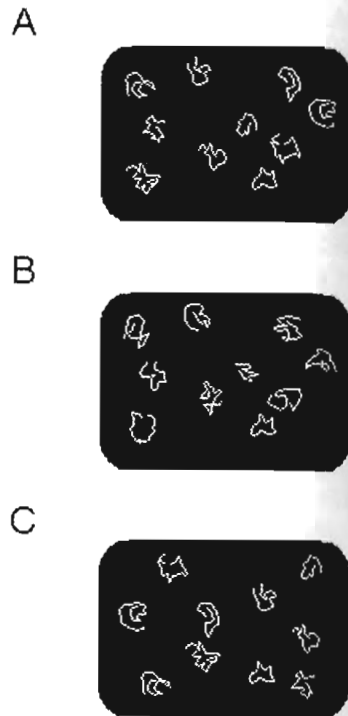
To convincingly demonstrate learning of spatial sequences, Mayr (1996; see also Willingham, Nissen, & Bullemer, 1989) employed a different task. Instead of responding to the spatial location, observers responded to the identity of each object as it appeared in different locations. In addition to the spatial sequence, the object identities formed their own sequence. The two sequences were uncorrelated, such that any effects of spatial sequence learning could be decoupled from response demands. Spatial sequence learning was demonstrated in a test block in which the sequence of spatial locations was scrambled and responses became slower; critically, because the identity sequence was preserved, this effect could not be attributed to response learning. As in the learning of spatial configurations, knowledge of the spatial sequences was implicit (Mayr, 1996). Subsequent research has demonstrated that static spatial configurations, in addition to spatial sequences, can be used to predict the location of a target that has yet to appear in the display—for example, the location of a visual search target on the next trial (Ono, Jiang, & Kawahara, 2005). All of these results reveal that we are remarkably sensitive to spatial statistical structure, whether conveyed in static or dynamic displays.

7.1.4.2 Spatial–object learning

We have seen how context distributed in time can facilitate target detection in space. Other examples demonstrate how information in two dimensions can get associated. As discussed earlier, spatial contextual cuing is a special situation, since the identities of the distractors are completely uninformative (beyond simply identifying them as distractors), and only the location of the target can be predicted. In this situation, observers can only learn associations between the spatial locations of the distractors and the target. In the visual environment, however, spatial configurations are often only as good as the objects they contain: The context of three objects arranged in a triangular configuration is not especially useful for finding a person’s mouth without knowing that these three objects are two eyes and a nose. Similarly in the laboratory, spatial contextual cuing can be attenuated when the distractor identities are altered, but the distractor locations are preserved (Jiang & Leung, 2005). This is not to say that spatial configurations alone provide no benefit, but rather that attaching diagnostic identity information can enhance learning and its subsequent expression.

To explore how the conjunction of spatial and object identity information affects search, Endo and Takeda (2004) employed another variant of the contextual cuing paradigm. Observers were presented with an array of outlined shapes, and were required to detect the shape whose contour was closed; the open-contour shapes served as distractors. Critically, each of the repeated target

Figure 7-6. Multidimensional contextual cuing. In Endo and Takeda (2004), observers performed a visual search task for the line drawing with the closed contour among distractors with open contours. In one experiment, the configuration of shapes was repeated during training, and, at test, observers search for the target in either (A) the original display (combined repetition), (B) a repetition of the configuration with new shapes (configuration repetition), or (C) a repetition of the shapes in a new spatial configuration (identity repetition). Only the combined and configuration repetitions facilitated visual search, suggesting that completely redundant identity information is discarded in contextual cuing. In another experiment in which identity repetitions were included in training (and thus, spatial configuration was not always reliable), all three types of repetition facilitated visual search at test.



and distractor locations was consistently mapped to a unique shape (Fig. 7-6A; *combined repetition*). Thus, both spatial configuration and distractor identity sets could predict both target location and target identity. In a test block, either the spatial configuration was left intact and the distractor identities randomized (Fig. 7-6B; *configuration repetition*), or the set of distractor identities was maintained and the locations randomized (Fig. 7-6C; *identity repetition*). Interestingly, the combined and configuration repetition facilitated target detection relative to a new configuration of shapes and locations, but the identity repetition did not. When all three types of old displays (combined, configuration, and identity repetition) were intermixed throughout training, rather than just being presented in the test block, target detection was facilitated by all three types of repetition. This surprising reversal suggests that, during visual search, completely redundant object context is not learned. Finally, in a different experiment in which spatial configuration predicted target identity and distractor identities predicted target location (the two untested combinations from Chun & Jiang, 1998, 1999), only the latter type of association was learned. Thus, regardless of the type of context, associations with target location are most beneficial during visual search.

Similar mechanisms may serve to associate objects and their most common spatial locations in a scene (Hollingworth, 2005). To explore this *object-in-place* learning, researchers can test how memory for individual objects depends on

whether they are presented in their original scene locations at test. For example, in one study (Hollingworth, 2006b), observers viewed a naturalistic scene for several seconds, and then completed a memory test. They were presented with one of the original objects from the scene and a modified version of the same object (rotated in depth, or a different exemplar of the same object type), and were required to determine which alternative was the original. Critically, although all test items were superimposed on the original background scene, the two alternatives were either presented in the same spatial location as the original, or in a new spatial location. Object memory—as defined by the accuracy of choosing the original object—was better when the alternatives were presented in the original location than in the new location, suggesting that the object's identity had been bound to that location.

Subsequent research has demonstrated that this form of learning is not strictly spatial; rather, it is relative to the configuration of the other objects in the display: Object memory is preserved in a new location if the configuration of the other objects is likewise translated (Hollingworth, 2007). Moreover, although such relational learning can improve explicit object recognition, it can also be observed using implicit eye movement measures. For example, when an object in a learned scene moves to a new location (thus violating the relational structure of the scene), it receives a disproportionate amount of viewing (e.g., Ryan, Althoff, Whitlow, & Cohen, 2000; Ryan, Leung, Turk-Browne, & Hasher, 2007). In fact, this increased sampling often only occurs when observers lack explicit awareness of the relational manipulation itself (Ryan et al., 2000).

NEURAL MECHANISMS FOR VISUAL ASSOCIATIVE LEARNING

Our review thus far describes the many rich ways in which the visual system encodes associations between objects and object parts across both space and time. The neural mechanisms that support such powerful learning are complex, yet great progress has been made toward understanding this highly adaptive and important function.

All brain systems are plastic (Gaffan, 2002). From the earliest stages of cortical processing in striate cortex to high-level representations in medial temporal and frontal cortices, such plasticity optimizes visual processing for features, objects, and scenes. This section will consider how visual associations may be encoded in the brain, which may eventually help us understand the organization of visual knowledge. We will first discuss basic principles for how visual associations can be represented in local neuronal circuitry. Then, we will discuss the neural substrates of visual associative learning at a systems level, describing how different structures in the brain may encode different types of visual associations. Specifically, we will focus on the medial temporal lobe system (MTL), which plays a central role in visual associative learning, but is still the subject of active debate about the specific functions of the different subregions, including the hippocampus, parahippocampal cortex, and perirhinal cortex.

7.2.1 Basic Mechanisms for Associative Learning

Miyashita and colleagues conducted classic studies to demonstrate how visual associations are encoded in temporal cortex (Miyashita, 1988; Sakai & Miyashita, 1991). In the first phase, they trained monkeys on visual shapes that were novel, avoiding contamination from preexisting associations that exist for familiar shapes. With such training, individual neurons in inferotemporal (IT) cortex become shape-selective (Kobatake, Wang, & Tanaka, 1998; Logothetis, Pauls, & Poggio, 1995; Mishkin, 1982). Inferotemporal cortex is the final stage of the ventral processing stream devoted to object vision, and it is where visual long-term representations are stored (Miyashita, 1993). Then, going beyond exhibiting such selectivity for individual trained stimuli, they demonstrated that IT neurons can also associate these geometrically unrelated items through co-occurrence over time (Miyashita, 1988; Sakai & Miyashita, 1991). In other words, if a neuron shows a good response to stimulus A, then it can also become responsive to other different-looking stimuli that appeared close in time to stimulus A during training. In theory, such mechanisms of associative learning in shape-selective neurons should not only allow for object-to-object priming between associated items, but also for the integration of different views of the same object (Logothetis & Pauls, 1995; Logothetis et al., 1995). More specifically in perirhinal cortex, a region in anterior medial IT cortex, Erickson and Desimone (1999) demonstrated that perirhinal neurons responded more similarly to associated stimuli, and significant activity occurred during the delay period between a predictor stimulus and its associated stimulus appearing after the delay.

Given such basic mechanisms of associative learning, one can ask about the functional organization of visual long-term memory. Cognitive models suggest that semantically related items should have strong, "close" connections, whereas unrelated items should have weaker, more "distant" connections. In the case of visual learning, this literally seems to be true. Erickson, Jagadeesh, and Desimone (2000) recorded from pairs of neurons in perirhinal cortex of macaques while the monkeys viewed series of complex visual images. When the stimuli were novel, neuronal responses for pairs of nearby neurons and far-apart neurons were uncorrelated. However, even after just 1 day of experience, the response preferences of nearby neurons became more similar, suggesting that learning induces the development of clusters with similar stimulus preferences. This finding is consistent with similar results obtained for simpler objects in more posterior regions of IT cortex (Gawne, Kjaer, Hertz, & Richmond, 1996; Gochin, Colombo, Dorfman, Gerstein, & Gross, 1994; Tanaka, 1996). Such evidence for local clustering makes it interesting to consider whether different temporal lobe brain regions may contain functional subdivisions at a more global level.

7.2.2 Basic Global Organization

The MTL system is critical for long-term storage of information. Key brain structures that we will focus on here include the hippocampus, parahippocampal cortex, and perirhinal cortex. Beyond the MTL proper, retrosplenial cortex appears to play a role in visual long-term associative memory as well.

A fundamental question is whether these diverse components of the MTL operate in a unitary manner or whether multiple functional subdivisions exist. Converging evidence suggests some degree of functional specialization. For example, the hippocampus is important for spatial processing, whereas perirhinal cortex is involved in object processing (for a review, see Murray, Graham, & Gaffan, 2005). Yet, it remains controversial whether hippocampal function should be segregated from other MTL structures using dichotomies such as associative versus nonassociative memory, declarative versus nondeclarative memory, explicit versus implicit memory, episodic versus semantic memory, and recollection versus familiarity (Squire, Stark, & Clark, 2004).

7.2.2.1 Hippocampus

The hippocampus performs a central function in associative and contextual learning. Contextual learning requires the binding of multiple cues. Thus, it can also be described as configural, spatioconfigural, or relational learning (Cohen & Eichenbaum, 1993; Hirsh, 1974; Johnson & Chalfonte, 1994; Kim & Fanselow, 1992; Moscovitch, 1994; Nadel & Moscovitch, 1997; O'Keefe & Nadel, 1978; Sutherland & Rudy, 1989). The Morris Water Maze task has been extensively studied to reveal the importance of the hippocampus in spatial contextual learning (Morris, Garrud, Rawlins, & O'Keefe, 1982). When rats are placed in a tank of opaque water, they can adeptly navigate to a hidden platform based on spatial cues provided by the room context. However, when the hippocampus is damaged, they lose the ability to perform this navigation task. Contextual fear conditioning is another popular measure of hippocampus-dependent configural learning (Kim & Fanselow, 1992; Phillips & LeDoux, 1992). When rats receive an aversive, electric shock in a cage, they become conditioned to associate the shock with the cage. When placed in the same cage after conditioning, they exhibit a fear-related freezing response. However, such fear conditioning is abolished with hippocampal ablation.

These findings suggest that the hippocampus and associated MTL structures should be important for visual context learning. As discussed earlier, one straightforward way to study visual context learning is to measure how subjects learn the positions of objects within scenes (Hollingworth, 2006a, 2006b). Gaffan, Murray, and colleagues demonstrated that hippocampus, perirhinal, and fornix lesions significantly impaired the ability to locate visual target shapes on complex backgrounds (Gaffan, 1994; Gaffan & Parker, 1996; Murray, Baxter, & Gaffan, 1998). Furthermore, electrophysiological recordings provide direct evidence that hippocampal neurons change their response properties during a learning task that required associating target locations with background scenes (Wirth et al., 2003). In human subjects tested with fMRI, an elegant study showed that the hippocampus and parahippocampal regions were sensitive to the bindings of objects with their backgrounds, showing fMRI adaptation only when the object-background pairings were repeated (Goh et al., 2004).

Neuropsychological studies provide further insight into the role of the hippocampus in contextual learning. Patients with hippocampal and extended MTL damage exhibited impairments in the spatial contextual cueing task

described earlier (Chun & Phelps, 1999). Contextual cueing can also be disrupted with neuropharmacological manipulations such as midazolam, which produces transient amnesia (Park, Quinlan, Thornton, & Reder, 2004). Importantly, the contextual learning in both of these two studies was implicit in all observers. Hence, these findings must be explained in terms of impaired spatial context learning rather than impaired conscious, explicit memory, which is also dependent on the hippocampus and MTL system (Squire, 1992).

However, the role of the hippocampus in implicit spatial associative learning is controversial. Manns and Squire (2001) compared contextual cueing in amnesic patients with restricted hippocampal damage and amnesic patients with more extended MTL damage. Only patients with extended MTL damage showed no spatial contextual cuing; patients with focal damage to the hippocampus (average 30% atrophy) showed normal contextual cuing effects. Because hippocampal damage was not complete in these patients, however, there remains some ambiguity about how these findings should be interpreted. Nevertheless, the patients were amnesic for declarative memory tasks, suggesting that the damage was consequential, and thus the dissociation between focal hippocampal and extended MTL damage deserves careful attention. Needless to say, human patient lesions are not precise enough to resolve refined questions of functional specialization.

Animal lesion studies and human brain imaging evidence using implicit tasks may provide further insight. For example, a brain imaging study revealed increased hippocampal and parahippocampal activation in tasks that required associating face stimuli and house stimuli, compared to single-item learning (Henke, Buck, Weber, & Wieser, 1997). Another study demonstrated hippocampal and parahippocampal involvement during associative encoding of line drawings of unrelated objects, for example, a picture of a monkey holding an umbrella (Rombouts et al., 1997). Thus, the hippocampus and posterior parahippocampal gyrus are involved in both the encoding and retrieval of novel picture pairs. Even for implicit contextual cueing tasks, fMRI reveals significant hippocampal activation (Greene et al., 2007). Although neuroimaging studies do not establish the necessity of a brain region for a given function, converging evidence from both human and animal models suggest that the hippocampus is important for both explicit and implicit learning of visual associations.

7.2.2.2 Parahippocampal cortex

The ability to perceive and represent one's local visual environment from scene information is fundamental for navigation and other visual behaviors. Thus, it is not surprising that dedicated neural machinery is specialized for processing scene information in medial temporal cortex (Aguirre, Detre, Alsup, & D'Esposito, 1996; Aguirre, Zarahn, & D'Esposito, 1998; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998a, 1998b). Termed the PPA, this region shows higher fMRI activation to scenes and three-dimensional spatial layouts than to faces, single objects, objects in arrays without three-dimensional layout information, or scrambled scene stimuli. Figure 7-7 shows individual PPA regions of five subjects tested in our lab. A robust PPA can be localized in

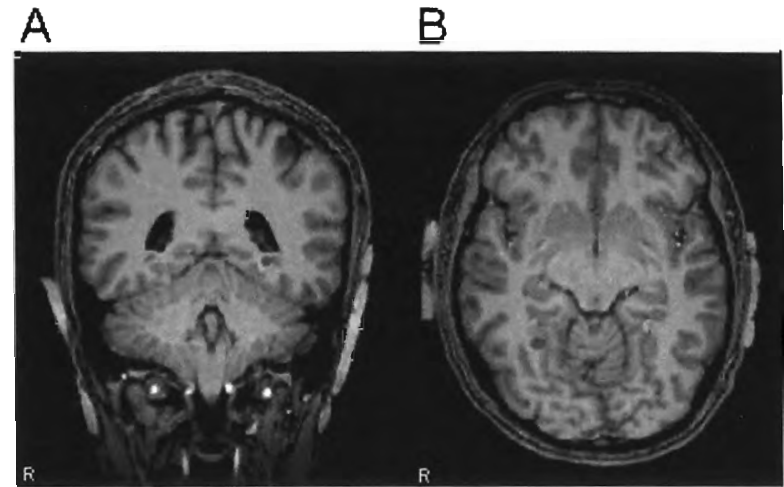


Figure 7-7. Bilateral parahippocampal place area (PPA) of five typical observers (in different colors) in coronal (A) and oblique axial (B) views. Regions-of-interest were defined as a 4-mm sphere surrounding the voxel on the collateral sulcus/parahippocampal gyrus showing the greatest difference between the hemodynamic response to scenes and faces.

almost all observers in both hemispheres, and the precise anatomical location is remarkably consistent across individuals.

It is quite interesting that the PPA is more responsive to spatial structure than to component objects per se. This is relevant because encoding spatial information requires configural, associative encoding, as shown earlier for the hippocampus. The PPA's sensitivity to spatial structure can be seen most clearly in Figure 7-8, which shows how this region responds to various types of stimuli (Epstein & Kanwisher, 1998a). Although the PPA response was most robust to scenes that contain rich structure, such as outdoor or indoor naturalistic scenes, it was also robust to even an empty room that did not contain any objects. Notably, the empty room response was higher than to a two-dimensional picture of furniture from a room on a blank background that lacked three-dimensional layout information.

However, the PPA does exhibit a significant response to objects as well, even when presented in isolation. The parahippocampal region is large in anatomical extent, beyond what is localized in the scene-object comparison used to define the PPA, and it is significantly involved in a wide variety of episodic memory tasks (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Turk-Browne, Yi, & Chun, 2006; Wagner et al., 1998). Thus, beyond the PPA's role in encoding spatial layout, we will review next how object associations are encoded in parahippocampal cortex.

One line of work provides clear evidence that the parahippocampal region is important for object contextual memory. Bar and Aminoff (2003) asked subjects to simply name a visually presented object on each trial. They compared

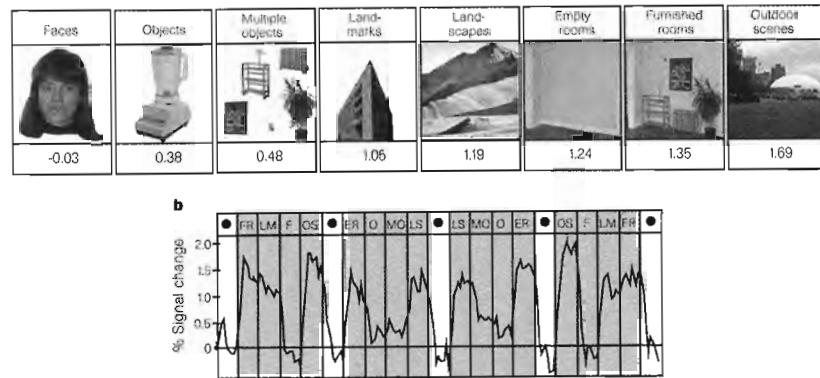


Figure 7-8. Example functional magnetic resonance imaging (fMRI) responses in the parahippocampal place area (PPA) to eight different stimulus types. Reprinted with permission from Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601. Macmillan Publishers Ltd.

cortical processing for two different sets of objects. Objects with strong contextual associations (strong CA objects) were rated to have strong associations with specific contexts, such as a shopping cart for supermarkets or microscopes for labs. In contrast, weak CA objects did not have such strong associations because they could occur in a wider variety of situations, such as a rope or a camera. The clever hypothesis was that pictures of strong CA objects should activate additional information associated with corresponding contexts, whereas weak CA objects should not elicit such contextual information. In other words, strong CA objects should trigger more activity in those brain regions that store or process visual contextual knowledge. Functional magnetic resonance imaging revealed two main sites of stronger activation for strong CA objects relative to weak CA objects. The first and largest focus was in the posterior part of the parahippocampal cortex, and the anatomical coordinates of this region (Talairach -24, -41, -4) overlap with those previously reported for the PPA. Another major region of activation was in the retrosplenial cortex, which is involved in episodic memory tasks and spatial navigation; retrosplenial function will be discussed in more detail in the next section.

Parahippocampal cortex and retrosplenial cortex are active to both spatial and nonspatial contexts. Interestingly, some gradation of spatial processing occurred in the parahippocampal gyrus, with stronger effects of spatial context along the posterior section and stronger effects of nonspatial context in the anterior section (Bar & Aminoff, 2003). The posterior end showed stronger overlap with the PPA (Epstein & Kanwisher, 1998a). This posterior–anterior axis of spatial specificity was confirmed even more directly in a nice study that directly compared spatial associations and nonspatial associations (Aminoff, Gronau, & Bar, 2006). Parahippocampal cortex was involved for both types of associative encoding, but nonspatial associations were mainly encoded in

anterior parahippocampal cortex, whereas spatial associations were represented in posterior parahippocampal cortex. Interestingly, only the latter posterior regions involved in spatial context learning overlapped with the PPA, defined by comparing scenes versus faces and scrambled scenes.

7.2.2.3 Retrosplenial cortex

In addition to the parahippocampal cortex, retrosplenial cortex is important for contextual associations, scene representation, and navigation (Aguirre & D'Esposito, 1999; Bar, 2004). Retrosplenial cortex runs as an arch around the splenium of the corpus callosum, encompassing the isthmus of the cingulate gyrus (Maguire, 2001). Navigation tasks represent an especially challenging function that requires the ability to associate and sequence disparate views and locations. Neuroimaging studies also point to a role for the retrosplenial cortex in spatial navigation tasks such as navigating through a virtual-reality maze (Aguirre et al., 1996) or when London taxi drivers were asked to recall complex routes around the city (Maguire, Frackowiak, & Frith, 1997). Beyond explicit navigation tasks, the role of retrosplenial cortex in associative processing is reinforced by several neuroimaging studies showing its involvement in episodic memory tasks (Andreasen et al., 1995; Fink et al., 1996; Wiggs, Weisberg, & Martin, 1999). Accordingly, retrosplenial damage produces general memory problems (Valenstein et al., 1987). Some interesting lateralization in function also appears to occur: Left retrosplenial cortex is more involved in general aspects of episodic memory, whereas right retrosplenial cortex is more involved in topographical orientation and spatial navigation (Maguire, 2001).

Interesting functional differences are apparent with parahippocampal cortex. Damage to retrosplenial cortex impairs the ability to orient oneself or navigate around the environment, even while the ability to identify scenes or landmarks remains intact (Maguire, 2001; Takahashi, Kawamura, Shiota, Kasahata, & Hirayama, 1997). In contrast, damage to parahippocampal cortex compromises the basic ability to identify scenes that lack individual distinctiveness (Mendez & Chierri, 2003). Thus, the neuropsychological evidence suggests that parahippocampal and retrosplenial cortex may play different roles in scene perception and memory.

Epstein and Higgins (2007) raised an interesting proposal to understand the functional differences between parahippocampal and retrosplenial cortex. They proposed that parahippocampal cortex is involved in identifying specific locations from the geometry and spatial cues in scenes, whereas the retrosplenial cortex plays a more integrative role in associating the scene with other related scenes to form a global topographic representation useful for directional navigation. In support of this hypothesis, they found that, compared to parahippocampal responses, retrosplenial activity was relatively weak when spatial scenes were merely viewed, but much stronger when task cues required subjects to retrieve information about the larger environment surrounding the scenes. In further support of an integrative function, responses in retrosplenial cortex, relative to parahippocampal cortex, are independent of exact physical properties (Bar & Aminoff, 2003).

Extending these ideas further, preliminary work in our lab explored the hypothesis that scene representations should be more viewpoint-invariant in retrosplenial cortex than in parahippocampal cortex (Park & Chun, under revision). Subjects were presented with three panoramic views of the same scene; each view overlapped with its adjacent view by only 30% of the image area. A viewpoint-dependent response would treat these three views as different from each other, whereas a viewpoint-independent response would treat these three views as similar to each other. Similarity can be measured by the fact that neural responses and corresponding fMRI signals are lower for sequences of similar-looking images than for different-looking images. In parahippocampal cortex, fMRI responses were similar for panoramic images as for sequences of novel images, suggesting that the panoramic images were treated as different from each other; that is, the responses were viewpoint-dependent. In contrast, responses to panoramic images were lower than to novel images in retrosplenial cortex, suggesting that these were treated as the "same," in support of a more integrative, viewpoint-invariant representation of the scenes from which the panoramic views were drawn. Future work will further clarify the differences and interactions between parahippocampal and retrosplenial processing.

7.2.2.4 Perirhinal cortex

Perirhinal cortex, which is located at the ventromedial aspect of the primate temporal lobe, is important for both the perception and memory of objects (Gaffan & Parker, 1996; Murray & Bussey, 1999). Perirhinal cortex is well suited to associate features of objects or objects with other objects, as it receives diffuse projections from inferotemporal cortex area TE (Saleem & Tanaka, 1996; Suzuki & Amaral, 1994). Murray and Richmond (2001) have summarized several functions of the perirhinal cortex.

First, perirhinal cortex is involved in object memory. Activity in perirhinal cortex neurons reflect long-term familiarity of visual images, as overall neuronal firing rate is higher to familiar relative to novel images (Holscher, Rolls, & Xiang, 2003). Also, ablation of perirhinal cortex impairs the ability to match an object with a target object presented previously (Buffalo et al., 1999; Meunier, Bachevalier, Mishkin, & Murray, 1993).

Second, perirhinal cortex is important for perception as well as memory. Perceptual factors can be tested by requiring observers to discriminate objects under various challenges, such as visual degradation or visual transformations. For example, recognition of complex objects is commonly viewpoint-dependent; that is, it becomes more difficult as the test object is rotated away from the original viewing angle. Monkeys with perirhinal damage have greater difficulty matching the same object presented in different viewing orientations (Buckley & Gaffan, 1998b), although human subjects with perirhinal damage do not show such impairments relative to control subjects with intact perirhinal cortex. Although perirhinal damage does not always produce impairments in difficult perceptual discriminations (Stark & Squire, 2000), it does seem to be important for difficult, ambiguous discriminations of complex objects; that is, when objects cannot be distinguished based on unique visual features alone

(Buckley, Booth, Rolls, & Gaffan, 2001; Lee et al., 2005; but see Levy, Shrager, & Squire, 2005).

Third, perirhinal cortex associates objects with other objects. Learning visual paired associates is impaired by perirhinal and entorhinal lesions (Buckley & Gaffan, 1998a; Higuchi & Miyashita, 1996; Miyashita, Okuno, Tokuyama, Ihara, & Nakajima, 1996; Murray, Gaffan, & Mishkin, 1993). Furthermore, brain-derived neurotrophic factor, which is involved in activity-dependent neural reorganization, is upregulated in perirhinal cortex during pair-associate learning (Tokuyama, Okuno, Hashimoto, Xin Li, & Miyashita, 2000). Perirhinal cortex also associates information about trial sequences. Monkeys readily learn visual cues that indicate progress toward a trial in which juice reward would be provided, but lesions to rhinal cortex impaired this ability to learn the predictive visual cues (Liu, Murray, & Richmond, 2000). Recent work is starting to distinguish whether perirhinal cortex is important for encoding, retrieval, or both stages of associative memory. An fMRI study suggests that perirhinal cortex is only active during encoding, but not retrieval of novel picture pairs, in contrast to the activation of hippocampus and parahippocampal gyrus areas in both encoding and retrieval (Pihlajamaki et al., 2003).

7.3 CONCLUSION

Visual objects and events always appear in a rich context of other objects and events, distributed over space and time. Statistical regularities occur in how objects appear with each other, and the brain has evolved exquisite mechanisms to encode such regularities, both within its local circuitry and at the systems level. Understanding these associative learning mechanisms will allow us to characterize the nature of visual knowledge and representation. As we aimed to highlight here, the effort to understand visual associative learning will require the continued development of innovative behavioral tasks, integrative theory, as well as advanced neuroscientific methods and techniques.

REFERENCES

- Aguirre, G. K., and D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain* 122(Pt 9):1613–1628.
- Aguirre, G. K., Detre, J. A., Alsup, D. C., and D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex* 6(6):823–829.
- Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). Neural components of topographical representation. *Proceedings of the National Academy of Sciences of the United States of America* 95(3):839–846.
- Aminoff, E., Gronau, N., and Bar, M. (2006). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex* 17(7):1493–1503.
- Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., et al. (1995). Remembering the past: Two facets of episodic memory explored with positron emission tomography. *American Journal of Psychiatry* 152(11):1576–1585.
- Avrami, J., and Kareev, Y. (1994). The emergence of events. *Cognition* 53(3):239–261.

- Baker, C. I., Olson, C. R., and Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science* 15(7):460–466.
- Bar, M. (2004). Visual objects in context. *Nat Rev Neurosci* 5(8):617–629.
- Bar, M., and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron* 38(2):347–358.
- Bar, M., and Ullman, S. (1995). *Spatial context in recognition*. Paper presented at the ARVO, Fort Lauderdale, FL.
- Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology* 14(2):143–177.
- Boyce, S. J., and Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:531–543.
- Boyce, S. J., Pollatsek, A., and Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance* 15:556–566.
- Brady, T. F., and Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance* 33:798–815.
- Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered [see comments]. *Science* 281(5380):1185–1187.
- Brockmole, J. R., Castelhano, M. S., and Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning Memory and Cognition* 32(4):699–706.
- Brockmole, J. R., and Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition* 13:99–108.
- Buckley, M. J., Booth, M. C., Rolls, E. T., and Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *Journal of Neuroscience* 21(24):9824–9836.
- Buckley, M. J., and Gaffan, D. (1998a). Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia* 36(6):535–546.
- Buckley, M. J., and Gaffan, D. (1998b). Perirhinal cortex ablation impairs visual object identification. *Journal of Neuroscience* 18(6):2268–2275.
- Buffalo, E. A., Ramus, S. J., Clark, R. E., Teng, E., Squire, L. R., and Zola, S. M. (1999). Dissociation between the effects of damage to perirhinal cortex and area te. *Learning and Memory* 6(6):572–599.
- Cavanagh, P., Labianca, A. T., and Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition* 80(1–2):47–60.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Science* 4(5):170–178.
- Chun, M. M. (2003). Scene perception and memory. In *Psychology of Learning and Motivation: Advances in Research and Theory: Cognitive Vision Vol. 42*, eds. D. Irwin and B. Ross, 79–108. San Diego: Academic Press.
- Chun, M. M., and Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology* 36:28–71.
- Chun, M. M., and Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science* 10:360–365.
- Chun, M. M., and Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning Memory and Cognition* 29:224–234.

- Chun, M. M., and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience* 2(9):844–847.
- Cohen, N. J., and Eichenbaum, H. (1993). *Memory Amnesia and the Hippocampal System*. Cambridge, MA: MIT Press.
- Cox, D., Meyers, E., and Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science* 304(5667):115–117.
- Davenport, J. L., and Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science* 15(8):559–564.
- Eckstein, M. P., Drescher, B. A., and Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and bayesian priors. *Psychological Science* 17(11):973–980.
- Egley, R., Driver, J., and Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General* 123(2):161–177.
- Endo, N., and Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception and Psychophysics* 66(2):293–302.
- Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron* 23(1):115–125.
- Epstein, R., and Kanwisher, N. (1998a). A cortical representation of the local visual environment. *Nature* 392(6676):598–601.
- Epstein, R., and Kanwisher, N. (1998b). The parahippocampal place area: A cortical representation of the local visual environment. *Journal of Cognitive Neuroscience*, 20–20.
- Epstein, R. A., and Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex* 17(7):1680–1693.
- Erickson, C. A., and Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience* 19(23):10404–10416.
- Erickson, C. A., Jagadeesh, B., and Desimone, R. (2000). Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nature Neuroscience* 3(11):1143–1148.
- Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., and Heiss, W. D. (1996). Cerebral representation of one's own past: Neural networks involved in autobiographical memory. *Journal of Neuroscience* 16(13):4275–4282.
- Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science* 12(6):499–504.
- Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning Memory and Cognition* 28(3):458–467.
- Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General* 134(4):521–537.
- Gaffan, D. (1994). Scene-specific memory for objects: A model of episodic memory impairment in monkeys with fornix transection. *Journal of Cognitive Neuroscience* 6:305–320.
- Gaffan, D. (2002). Against memory systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357(1424):1111–1121.

- Gaffan, D., and Parker, A. (1996). Interaction of perirhinal cortex with the fornix-fimbria: Memory for objects and "object-in-place" memory. *Journal of Neuroscience* 16:5864–5869.
- Gawne, T. J., Kjaer, T. W., Hertz, J. A., and Richmond, B. J. (1996). Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cerebral Cortex* 6(3):482–489.
- Gibson, E. J. (1963). Perceptual learning. *Annual Review of Psychology* 14:29–56.
- Gibson, E. J. (1966a). *Perceptual development and the reduction of uncertainty*. Paper presented at the Proceedings of the 18th International Congress of Psychology, Moscow.
- Gibson, J. J. (1966b). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L., and Gross, C. G. (1994). Neural ensemble coding in inferior temporal cortex. *Journal of Neurophysiology* 71(6):2325–2337.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., and Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition* 12(6):878–892.
- Goh, J. O., Siong, S. C., Park, D., Gutchess, A., Hebrank, A., and Chee, M. W. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *Journal of Neuroscience* 24(45):10223–10228.
- Greene, A. J., Gross, W. L., Elsingher, C. L., Rao, S. M. (2007). Hippocampal differentiation without recognition: An fMRI analysis of the contextual cueing task. *Learning & Memory*, 14, 548–553.
- Harris, Z. (1955). From phoneme to morpheme. *Language: Journal of the Linguistic Society of America* 31:190–222.
- Henderson, J. M., and Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology* 50:243–271.
- Henke, K., Buck, A., Weber, B., and Wieser, H. G. (1997). Human hippocampus establishes associations in memory. *Hippocampus* 7(3):249–256.
- Higuchi, S., and Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Sciences of the United States of America* 93(2):739–743.
- Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioral Biology* 12(4):421–444.
- Hollingworth, A. (2005). Memory for object position in natural scenes. *Visual Cognition* 12(6):1003–1016.
- Hollingworth, A. (2006a). Memory for object position in natural scenes. *Visual Cognition* 12:1003–1016.
- Hollingworth, A. (2006b). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning Memory and Cognition* 32(1):58–69.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance* 33, 31–47.
- Hollingworth, A., and Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General* 127:398–415.
- Holscher, C., Rolls, E. T., and Xiang, J. (2003). Perirhinal cortex neuronal activity related to long-term familiarity memory in the macaque. *European Journal of Neuroscience* 18(7):2037–2046.

- Howard, J. H., Jr., Dennis, N. A., Howard, D. V., Yankovich, H., and Vaidya, C. J. (2004). Implicit spatial contextual learning in healthy aging. *Neuropsychology* 18:124–134.
- Hunt, R. H., and 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(4):658–680.
- Jiang, Y., and Leung, A. W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin and Review* 12:100–106.
- Jiang, Y., and Song, J-H (2005). Hyper-specificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception and Performance* 31(6):1439–1448.
- Jiang, Y., Song, J-H, and Rigas, A (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin and Review* 12(3):524–529.
- Jiang, Y., and Wagner, L. C. (2004). What is learned in spatial contextual cueing: Configuration or individual locations? *Perception and Psychophysics* 66:454–463.
- Jiménez, L., and Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning Memory and Cognition* 25:236–259.
- Johnson, J. S., Woodman, G. F., Braun, E., and Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin and Review* 14:834–837.
- Johnson, M. K., and Chalfonte, B. L. (1994). Binding complex memories: The role of reactivation and the hippocampus. In *Memory Systems*, eds. D. L. Schacter and E. Tulving, 311–350. Cambridge, MA: MIT Press.
- Kawahara, J. (2003). Contextual cueing in 3-D layouts defined by binocular disparity. *Visual Cognition* 10:837–852.
- Kim, J. J., and Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science* 256(5057):675–677.
- Kobatake, E., Wang, G., and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology* 80(1):324–330.
- Kunar, M. A., Flusberg, S., Horowitz, T. S., and Wolfe, J. M. (in press). Does contextual cueing guide the deployment of attention. *Journal of Experimental Psychology: Human Perception and Performance* 33:816–828.
- Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., et al. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia* 43(1):1–11.
- Levy, D. A., Shrager, Y., and Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learning and Memory* 12(1):61–66.
- Liu, Z., Murray, E. A., and Richmond, B. J. (2000). Learning motivational significance of visual cues for reward schedules requires rhinal cortex. *Nature Neuroscience* 3(12):1307–1315.
- Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance* 4:565–572.
- Logothetis, N. K., and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex* 5(3):270–288.
- Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology* 5(5):552–563.

- Luck, S. J., Girelli, M., McDermott, M. T., and Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology* 33:64–87.
- Mackworth, N. H., and Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics* 2:547–552.
- Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology* 42(3):225–238.
- Maguire, E. A., Frackowiak, R. S., and Frith, C. D. (1997). Recalling routes around London: Activation of the right hippocampus in taxi drivers. *Journal of Neuroscience* 17(18):7103–7110.
- Manns, J., and Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus* 11:776–782.
- Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences* 6(6):255–260.
- 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(2):350–364.
- McCotter, M., Gosselin, F., Sowden, P., and Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition* 12(6):938–953.
- Mendez, M. F., and Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia* 41(10):1387–1395.
- Meunier, M., Bachevalier, J., Mishkin, M., and Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *Journal of Neuroscience* 13(12):5418–5432.
- Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 298(1089):83–95.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335(6193):817–820.
- Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience* 16:245–263.
- Miyashita, Y., Okuno, H., Tokuyama, W., Ihara, T., and Nakajima, K. (1996). Feedback signal from medial temporal lobe mediates visual associate mnemonic codes of inferotemporal neurons. *Cognitive Brain Research* 5:81–86.
- Morris, R. G., Garrud, P., Rawlins, J. N., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature* 297(5868):681–683.
- Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In *Memory Systems*, eds. D. L. Schacter and F. Tulving, 269–310. Cambridge, MA: MIT Press.
- Murray, E. A., Baxter, M. G., and Gaffan, D. (1998). Monkeys with rhinal cortex damage or neurotoxic hippocampal lesions are impaired on spatial scene learning and object reversals. *Behavioral Neuroscience* 112(6):1291–1303.
- Murray, E. A., and Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends in Cognitive Sciences* 3:142–151.
- Murray, E. A., Gaffan, D., and Mishkin, M. (1993). Neural substrates of visual stimulus-stimulus association in rhesus monkeys. *Journal of Neuroscience* 13(10):4549–4561.
- Murray, E. A., Graham, K. S., and Gaffan, D. (2005). Perirhinal cortex and its neighbours in the medial temporal lobe: Contributions to memory and perception.

- Quarterly Journal of Experimental Psychology. B Comparative and Physiological Psychology* 58(3–4):378–396.
- Murray, E. A., and Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology* 11(2):188–193.
- Nadel, L., and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology* 7(2):217–227.
- Naor-Raz, G., Tarr, M. J., and Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception* 32(6):667–680.
- Neider, M. B., and Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research* 46(5):614–621.
- Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology* 19:1–32.
- O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
- Olson, I. R., and Chun, M. M. (2001). Temporal contextual cueing of visual attention. *Journal of Experimental Psychology: Learning Memory and Cognition* 27:1299–1313.
- Olson, I. R., and Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition* 9:273–302.
- Ono, F., Jiang, Y., and Kawahara, J. (2005). Intertrial temporal contextual cuing: Association across successive visual search trials guides spatial attention. *Journal of Experimental Psychology: Human Perception and Performance* 31:703–712.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition* 3:519–526.
- Park, H., Quinlan, J., Thornton, E., and Reder, L. M. (2004). The effect of midazolam on visual search: Implications for understanding amnesia. *Proceedings of the National Academy of Sciences of the United States of America* 101(51):17879–17883.
- Park, S., and Chun, M. M. (under revision). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception.
- Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences* 10(5):233–238.
- Peterson, M. S., and Kramer, A. F. (2001a). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception and Psychophysics* 63(7):1239–1249.
- Peterson, M. S., and Kramer, A. F. (2001b). Contextual cueing reduces interference from task-irrelevant onset distractors. *Visual Cognition* 8:843–859.
- Phillips, R. G., and LeDoux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral Neuroscience* 106(2):274–285.
- Pihlajamaki, M., Tanila, H., Hanninen, T., Kononen, M., Mikkonen, M., Jalkanen, V., et al. (2003). Encoding of novel picture pairs activates the perirhinal cortex: An fMRI study. *Hippocampus* 13(1):67–80.
- Posner, M. I., Snyder, C. R. R. and Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General* 109:160–174.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6:855–863.
- Rombouts, S. A., Machielsen, W. C., Witter, M. P., Barkhof, F., Lindeboom, J., and Scheltens, P. (1997). Visual association encoding activates the medial temporal lobe: A functional magnetic resonance imaging study. *Hippocampus* 7(6):594–601.

- Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science* 11(6):454–461.
- Ryan, J. D., Leung, G., Turk-Browne, N. B., and Hasher, L. (2007). Assessment of age-related changes in inhibition and binding using eye movement monitoring. *Psychology and Aging* 22:239–250.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition* 81(2):149–169.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants [see comments]. *Science* 274(5294):1926–1928.
- Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates [see comments]. *Nature* 354(6349):152–155.
- Saleem, K. S., and Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area to the perirhinal and entorhinal cortices in the macaque monkey. *Journal of Neuroscience* 16(15):4757–4775.
- Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review* 84(1):1–66.
- Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science* 5:195–200.
- Sheinberg, D. L., and Logothetis, N. K. (1998). Implicit memory for scenes guides visual exploration in monkey. *Society for Neuroscience Abstracts* 24(2):1506.
- Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84(2):127–190.
- Shin, J. C., and Ivry, R. B. (2002). Concurrent learning of temporal and spatial sequences. *Journal of Experimental Psychology: Learning Memory and Cognition* 28(3):445–457.
- Shomstein, S., and Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting. *Perception and Psychophysics* 64(1):41–51.
- Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience* 99:195–231.
- Squire, L. R., Stark, C. E., and Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience* 27:279–306.
- Stark, C. E., and Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learning and Memory* 7(5):273–278.
- Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., and Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron* 49(6):905–916.
- Sutherland, R. J., and Rudy, J. W. (1989). Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology* 17:129–144.
- Suzuki, W. A., and Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *Journal of Comparative Neurology* 350(4):497–533.
- Takahashi, N., Kawamura, M., Shiota, J., Kasahata, N., and Hirayama, K. (1997). Pure topographic disorientation due to right retrosplenial lesion. *Neurology* 49(2):464–469.
- Tanaka, J. W., and Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception and Psychophysics* 61(6):1140–1153.

- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19:109–139.
- Tokuyama, W., Okuno, H., Hashimoto, T., Xin Li, Y., and Miyashita, Y. (2000). BDNF upregulation during declarative memory formation in monkey inferior temporal cortex. *Nature Neuroscience* 3(11):1134–1142.
- Torralba, A., Oliva, A., Castelano, M. S., and 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(4):766–786.
- Turk-Browne, N. B., Isola, P. J., Scholl, B. J., and Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning Memory and Cognition* 24(2):399–407.
- Turk-Browne, N. B., Jungé, J., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General* 134(4):552–564.
- Turk-Browne, N. B., and Scholl, B. J. (in press). Flexible visual learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*.
- Turk-Browne, N. B., Yi, D.-J., and Chun, M. M. (2006). Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron* 49:917–927.
- Valenstein, E., Bowers, D., Verfaellie, M., Heilman, K. M., Day, A., and Watson, R. T. (1987). Retrosplenial amnesia. *Brain* 110(6):1631–1646.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281(5380):1188–1191.
- Wiggs, C. L., Weisberg, J., and Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia* 37(1):103–118.
- Willingham, D. B., Nissen, M. J., and Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition* 15(6):1047–1060.
- Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., and Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science* 300(5625):1578–1581.
- Zacks, J. M., and Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin* 127(1):3–21.