Obligatory and adaptive averaging in visual short-term memory

Chad Dubé

University of South Florida, City, FL, USA ✉

Robert Sekuler

Volen Center for Complex Systems, Brandeis University, Waltham, MA, USA ✉

Visual memory can draw upon averaged perceptual representations, a dependence that could be both adaptive and obligatory. In support of this idea, we review a wide range of evidence, including findings from our own lab. This evidence shows that time- and space-averaged memory representations influence detection and recognition responses, and do so without instruction to compute or report an average. Some of the work reviewed exploits fine-grained measures of retrieval from visual short-term memory to closely track the influence of stored averages on recall and recognition of briefly presented visual textures. Results show that reliance on perceptual averages is greatest when memory resources are taxed or when subjects are uncertain about the fidelity of their memory representation. We relate these findings to models of how summary statistics impact visual short-term memory, and discuss a neural signature for contexts in which perceptual averaging exerts maximal influence.

Introduction

In its broadest sense, a representation may be defined as anything that stands for something other than itself (Frisby & Stone, 2010). For example, the word *quinoa* is a representation of the grain quinoa, and if you are staring intently at some quinoa, the activations of neurons in cortical area V1 would be yet another representation of quinoa. In fact, the visual system exploits multiple representations, which vary in the fidelity with which each captures the details of the stimulus being represented. At one extreme, such representations may be detailed and precise, faithfully capturing a great many of a stimulus's features; at the other extreme, they may be likened to a broad-brush, quick sketch of the stimulus, which omits most details.

It is easy to imagine the value of discarding some sensory information in favor of a more compact, less detailed representation. For example, the spatial or temporal properties of some stimulus could limit the information that can be encoded, thereby forcing the system to fall back on a space- or time-averaged summary of the incoming stimulus. McDermott, Schemitsch, and Simoncelli (2013) made this point with particular clarity in a study of audition. Working with complex sounds whose features varied rapidly over time, they found that in many conditions, the auditory system's representation was a time-averaged mean of the stimulus stream. Basically, so much information had been presented so quickly that it outstripped the system's encoding and buffering capacity. Earlier, with rapidly presented visual sequences as well as auditory ones, Pollack (1974) noted that even when individual items in a sequence were imperceptible, the sequence's overall structure produced a distinctive percept. The availability of a compact representation affords one other important functional advantage: It facilitates communication of what is being heard or seen. Thus, under most circumstances, it is more useful to be told that that you are looking at a bowl of cooked quinoa than to be told you are looking at a bowl of 5,012 flat, discoid reddish seeds, each about 1.5 mm in diameter. Clearly, then, unnecessary detail can interfere with communication.

Vision researchers have demonstrated averaging effects in a variety of arenas. For example, such effects are seen in experiments with random-dot cinematograms, which comprise hundreds of spatially inter-mingled elements that move independently of one another. In many conditions, such stimuli produce percepts whose directional flow approximates a mean taken over all the directions that were present in the cinematogram, or whose flow speed approximates a mean taken over the speeds of the cinematogram's elements (Watamaniuk & Duchon, 1992; Watamaniuk & Sekuler, 1992; Watamaniuk, Sekuler, & Williams, 1989).

As Haberman and Whitney (2012) have noted, the idea that the visual system extracts summary statistics at the expense of individual features is far from a new one, going back at least as far as the writings of first-generation Gestalt psychologists (e.g., Koffka, 1935, pp. 270, 273). Despite the idea's long history and the recent increase in efforts to understand the statistical representations generated by the visual system, the basic structural mechanisms and functional significance of such summary statistics remain unclear. In what follows, we summarize some studies that suggest perceptual averaging is an adaptive mechanism, which has as one of its functions buffering the visual memory system against interference and overload. Our review focuses primarily on remembered averages that impact visual short-term memory responses rather than on averaging effects that occur under conditions of low memory load.[1] We also examine the implications of perceptual averaging for models of visual short-term memory in which recognition memory decisions are assumed to arise from comparisons of individual items' features. Finally, we describe recent data that point to a neural signature for the contexts in which the influence of perceptual averaging is most pronounced.

Space-time perceptual averaging

A study by Ariely (2001) embodies a central thread running through much research on perceptual averaging in visual short-term memory (VSTM). In that study, subjects were presented with memory sets comprising multiple circles whose diameters differed. Following this, a test stimulus with one or two probe circles was presented. The number of probes differed only with the nature of the test: yes/no (one probe) or two-alternative forced choice (two probes). Since results from the two tasks were similar, we will describe only the yes/no variant. Subjects judged the test stimulus in one of two ways. On some trials, they made a yes/no response as to whether the probe had been in the memory set (member identification). On other trials, the response indicated whether the probe circle was larger or smaller than the mean of the items' sizes in the memory set (mean discrimination). The results showed that despite mere chance performance on member identification, mean discrimination was quite good. On the basis of these results, Ariely speculated that "when presented with a set of four or more similar items, the visual system creates a representation of the set, and discards information about the individual items in the set" (p. 160). Furthermore, he suggested that the use of such statistics was adaptive: "The reduction of a set of similar items to a mean (or prototypical value), a range, and a few other important

statistical properties may preserve just the information needed to navigate in the real world, to form a stable global percept, and to identify candidate locations of interest" (p. 161).

A major contribution to work on perceptual averaging was the establishment of attention's role in these phenomena. Studies of attention's impact have supported Ariely's conclusion that perceptual averages may be preserved even in the absence of memory for individual items themselves (Ariely, 2001; Corbett & Oriet, 2011). For instance, Alvarez and Oliva (2008) presented subjects with sets of eight moving circles. The subjects' task was to count the number of times four of the moving circles (the target set) touched two red lines contained in the display. While making this judgment, subjects were to ignore the other four moving circles (the distracter set). Following this, a display was presented that contained either all but one of the circles, in their resting positions (individual test), or four of them in their resting positions (centroid test). In the individual test, subjects indicated the location that would have been occupied by the one missing circle. In the centroid test, they indicated the mean location of the four missing items. Alvarez and Oliva found that, though accuracy was higher for targets than distracters in the individual test, performance on the centroid test was equivalent for targets and distracters. This suggests that perceptual averages are preserved even for items that are filtered out of awareness by selective attention (see also Alvarez & Oliva, 2009).

Other studies of VSTM have used various tasks to explore different aspects of this averaging process. These tasks include motion detection (Ball & Sekuler, 1980), multiple-object tracking (Alvarez & Oliva, 2008), change detection (Alvarez & Oliva, 2009), rapid serial visual presentation (Corbett & Oriet, 2011), and Sternberg's memory scanning task (Dubé, Zhou, Kahana, & Sekuler, 2014), in addition to tasks in which subjects had to report some average feature of a briefly presented stimulus display (Chong & Treisman, 2005a, 2005b; Emmanouil & Treisman, 2008). These studies make it clear that perceptual averaging can operate on remembered direction, location, size, speed, texture, and even facial expression (Haberman, Harp, & Whitney, 2009).

Furthermore, perceptual averaging is not limited to stimuli in which elements are distributed over space, as in Ariely's and so many other studies. Perceptual averaging can also operate on elements that are distributed in time (e.g., Albrecht & Scholl, 2010; Corbett & Oriet, 2011; Dubé et al., 2014; Haberman et al., 2009). In one such study, Haberman et al. (2009) presented subjects with a series of briefly presented face morphs that varied in emotional expression. Subjects were asked whether a probe face was more or less disgusted-looking than the average expression of the

faces they had just seen. The results showed that subjects' judgments were quite accurate, demonstrating that perceptual averaging can occur across trials. Subsequent studies have extended this finding to paradigms involving abstract geometric stimuli of the sort commonly used in studies of visual search and change detection, as well as stimuli that change dynamically over time (e.g., Albrecht & Scholl, 2010). These important findings link perceptual averaging to earlier studies that suggest that average or prototypical features can be computed over successive trials in tasks involving comparative judgment (Morgan, Watamaniuk, & McKee, 2000), categorization (Busemeyer & Myung, 1988), and VSTM (Wilken & Ma, 2004). In other words, it seems that perceptual averages can be extracted from spatially defined sets of items or from temporally defined ones. This latter, temporal mode of averaging is important in part because it reminds us that an averaging process need not be restricted to events that occur within the experimenter-defined boundaries of an experimental trial (e.g., Morgan et al., 2000). Instead, temporal averaging can and does extend to events and items that span such boundaries.

In the aggregate, these studies suggest that perceptual averaging may be integral to visual memory, and that it can interact with and influence functions such as feature matching, object updating, expectation-based monitoring, and perceptual binding (Alvarez, 2011; Treisman, 2006). Although the details of experimental tasks and stimuli vary considerably among these studies, all have in common with Ariely's study the requirement that subjects compute a perceptual average and communicate some direct or indirect indicator of that computation. On the basis of such studies alone, it is difficult to say whether the averaging mechanism is essential to the visual memory system or merely activated and driven by the instruction to compute such an average. Recent work from our laboratory (discussed later) suggests that perceptual averaging is in fact an essential, obligatory aspect of memory encoding, one that influences VSTM responses even in the absence of any instruction to compute or report an average.

Obligatory averaging

Dating back more than three decades, investigations in detection and psychophysics have demonstrated perceptual averaging at work under conditions in which subjects were not explicitly asked to compute an average. In one such study (Ball & Sekuler, 1980), subjects responded to onset of motion in a random-dot cinematogram (RDC). In a series of experiments, the direction of motion was either certain (e.g., motion was
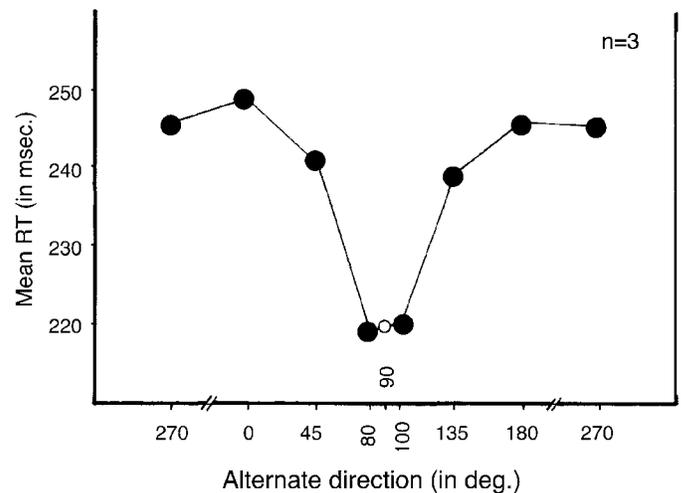

Figure 1. Mean RT to detect onset of upward (90°) motion. The open data point is the mean RT when only motion upward could occur (directional certainty). Each filled data point comes from a condition in which motion could, with equal probability, be either 90° or in the alternate direction shown on the horizontal axis. Adapted from "Models of Stimulus Uncertainty in Motion Perception," by K. Ball and R. Sekuler, 1980, *Psychological Review, 87,* p. 442. Copyright 1980 by the American Psychological Association. Adapted with permission.

always 90° upward) or uncertain (e.g., motion could be either 90° upward or 180° leftward horizontal, directions that varied quasi-randomly over trials). As shown in Figure 1, simple reaction times (RTs) showed a U-shaped pattern, increasing as the two RDC directions that were possible on any trial diverged from one another.

Detection of an RDC's direction requires subjects to compute an average over the constituent motion vectors making up the RDC (Watamaniuk & Sekuler, 1992). The RT pattern obtained across trials, however, suggests the operation of a complementary spatiotemporal process. Specifically, Ball and Sekuler tested the idea that when faced with uncertainty about which of two directions of motion might occur, subjects could elect to monitor visual channels maximally responsive to the direction midway between the two alternative directions. In one such test, subjects were presented with three possible directions of motion: 60°, 120°, and 90°. Here, the 60° and 120° RDC trials were presented equally often. The 90° RDC was an infrequent probe stimulus (presented on only 5% to 10% of trials) whose direction matched the perceptual average of the two directions. Such an infrequent stimulus would normally produce inflated RTs (Hyman, 1953). However, Ball and Sekuler found exactly the opposite: RTs were shortest to the infrequent probe RDC and did not vary with the separation between the two dominant directions of motion. This result supported the researchers' hypothesis that a representation had been encoded and

maintained over trials of the direction that lay midway between the two remembered RDC directions. They hypothesized that such representations could have been adopted in order to mitigate uncertainty about the stimulus that might be presented on a given trial.

Research with simple visual textures has gone further, suggesting that perceptual averaging is a basic and obligatory aspect of memory encoding. This fact is far from self-evident in most studies of perceptual averaging, which (following Ariely, 2001) have usually required that subjects make an explicit report of the central tendency of a set or sequence of stimuli. Naturally, requiring such a report encourages subjects to compute such a measure. Would such a statistic be computed without such pressure? Adapting a procedure developed by Huang and Sekuler (2010b), Dubé et al. (2014) recently answered that question in a study of stimulus uncertainty that did not require subjects to report an average. On each trial in that study, subjects saw two Gabor patches (sinusoidally modulated, windowed luminance gratings) presented sequentially. Each Gabor comprised both vertically and horizontally oriented sinusoidal components. Following the second Gabor, a recognition probe was presented that either did or did not match one of the two study Gabors (deemed the relevant study item, which had been cued in advance, as explained in the next paragraph). Subjects then made a yes/no recognition response to the probe. To this basic design, two key manipulations were added.

One key manipulation entailed the presentation of the attention-directing cue just mentioned. This cue was inserted several times during a trial. This made it possible to examine how selective attention affected short-term memory for the stimuli. The cue designated the stimulus (1 or 2) that would be that trial's relevant study item. In other words, the cue identified the stimulus whose spatial frequency subjects should compare to the probe item's spatial frequency. The attention-directing cue appeared either before the first study item (precue condition), between the two items (midcue), or after both items had been presented (postcue).

The second key manipulation entailed the level of similarity between the probe and the study items. The recognition probes presented to each subject in the experiment took on 15 different levels of similarity to the relevant study item (measured in just-noticeable differences (JNDs) of vertical spatial frequency), including the value that matched the relevant study item (target trials, having 4 or 12 JNDs of spatial frequency) and 14 degrees of mismatch (lure trials). Relevant and irrelevant study items were always separated by 8 JNDs, with the relevant item taking on a value of either 4 or 12 JNDs. With this technique, the distributions of recognition response rates—P("Old")

values—at each of several levels of feature matching could be constructed. These distributions, termed "mnemometric functions" (Sekuler & Kahana, 2007), were then modeled using a truncated skew-normal distribution (Azzalini, 1986). The outcome produced separate estimates of the Gaussian variance and skew of the resulting response distributions.

The results are shown in Figure 2. In the figure, the spatial frequency matching the relevant study item is scaled to 4 JNDs on the x-axis, and that of the irrelevant study item to 12 JNDs, in order to facilitate visual inspection.[2] Included are data from a baseline condition, called Single, in which only one study item was presented. The data show that perceptual averaging influenced responses to the probe stimulus. Specifically, subjects' response distributions around the target probe frequency (4 JNDs in the figure) were inflated in the region spanning the average of the two study items' spatial frequencies (5–11 JNDs). This occurred regardless of whether the first (panel A) or second (panel B) study item was the task-relevant item. Crucially, the analysis of the skew-normal parameters revealed a greater influence of perceptual averaging when the attention cue could not be used to selectively attend in advance of the relevant stimulus. This effect can be seen in the postcue conditions of Figure 2, labeled "post-1" and "post-2" (depending on whether the first or second study item was relevant). These postcue conditions should have produced maximal uncertainty about the fidelity of a given memory representation. The researchers concluded that perceptual averaging is a basic, perhaps obligatory operation for the visual memory system.

Recent work by Corbett and colleagues (Corbett & Melcher, 2014; Corbett & Song, 2014; Corbett, Wurnitsch, Schwartz, & Whitney, 2012) has made clever use of adaptation to track the effects of perceptual averaging in the absence of instruction. In one of these studies (Corbett & Melcher, 2014), subjects were presented with patches of stationary dots differing in diameter. Dots appeared in two clusters, one to the left and one to the right of fixation, and the clusters differed in the average diameter of the dots making up the cluster. This display served as an adapting stimulus, remaining onscreen for 1 min. Following this, two test clusters of dots replaced the previously seen clusters. Subjects were asked to indicate which test cluster's circles had the larger mean diameter. Subjects' psychometric functions showed that test clusters preceded by ensembles with a larger mean diameter appeared to have a smaller mean diameter, and those preceded by ensembles with a smaller mean diameter appeared to have a larger mean diameter. Interestingly these effects were preserved despite variation in the test dots' retinal locations, suggesting that the effects were not retinotopic. Like the study by Ball and Sekuler (1980), this
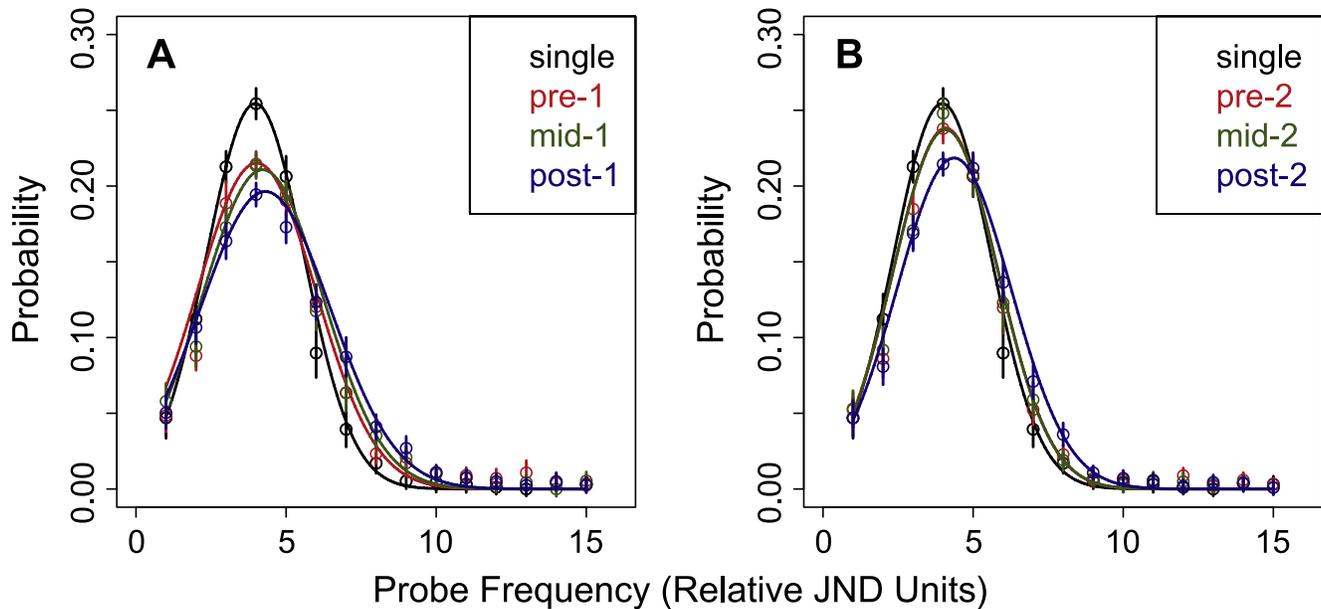
Figure 2. Mnemometric distributions showing elevated response probabilities to recognition probe Gabors whose spatial frequencies lie in the region (5–11 JNDs) spanning the average spatial-frequency feature value of two study Gabors held in memory. (A) Three selective attention conditions and a baseline (Single), when the first study item was task relevant. (B) The same conditions, but for trials on which the task-relevant item was the second study item. Reprinted from "Similarity-Based Distortion of Visual Short-Term Memory Is Due to Perceptual Averaging" by C. Dubé, F. Zhou, M. Kahana, and R. Sekuler, 2014, Vision Research, 96, p. 13. Copyright 2014 by Elsevier Ltd. Reprinted with permission.

work involves what is nominally a visual perception task. However, it provides another clear demonstration of obligatory averaging in memory.

Prototypes and exemplars in VSTM

The prevalence of perceptual averaging effects, even in the absence of explicit instruction, has theoretical importance for researchers interested in perceptual categorization (PC). Early studies of categorization demonstrated that subjects can extract and remember averaged or prototypical stimuli even when no single exemplar matching the prototype has been presented (Posner & Keele, 1968). Although such prototypes are often assumed to be computed over relatively large numbers of trials (Morgan et al., 2000; Murphy, 2002), it is conceivable that representations averaged over even a subset of recently encountered items could influence decisions in categorization just as they do in VSTM (Busemeyer & Myung, 1988; Gorea & Sagi, 2000; Huang & Sekuler, 2010b). Nonetheless, the most successful models for PC have arguably been ones that assume that categorization responses depend upon individual exemplars rather than prototypes computed over individual exemplars (Nosofsky, 1992). Importantly, exemplar models have recently been extended to VSTM, implying that individual items' features alone

determine VSTM responses. In this section, we will describe the exemplar approach to categorization, the assumptions of exemplar models, and some recent extensions to VSTM.

To understand PC, consider a simple scenario in which subjects study sets of visual stimuli (e.g., circles) which differ in one feature (e.g., the orientation of a line drawn along each circle's diameter). Suppose that, unbeknownst to subjects, each stimulus is drawn from one of two distributions of stimuli. One set of stimuli, Category A, has a distribution of orientations centered at 90° (upward), and the other, Category B, has a distribution centered at 180° (horizontal). Examples of such stimuli are displayed in Figure 3. Subjects must make a classification response (A or B) for each stimulus, and are provided feedback after each trial, allowing them to learn the category assignments over several trials. One way of modeling subjects' performance in this task is to assume that they extract an average over some or all of the exemplars shown over trials for a given class (as specified by feedback learning), and use these remembered prototypes as standards to which new exemplars are compared (Busemeyer & Myung, 1988; Murphy, 2002; Posner & Keele, 1968).

However, a number of classic findings in the PC literature argue strongly against prototype models and in favor of exemplar models of PC (for a review, see Murphy, 2002). Perhaps the most influential of these
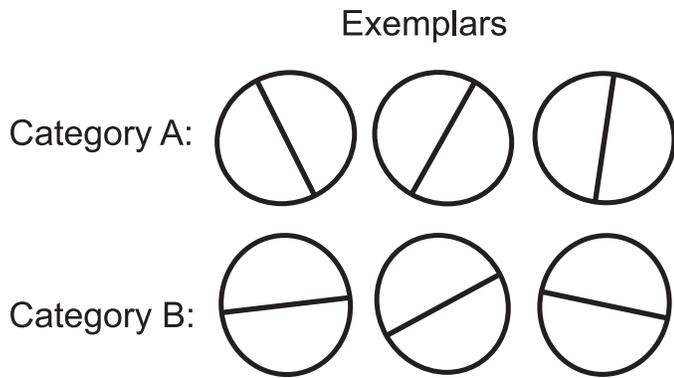
Exemplars

Figure 3. Example of a set of stimuli from a simple perceptual categorization task (e.g., Maddox & Ashby, 1993). Stimuli drawn from Category A are from a distribution of stimuli with a mean orientation that is vertical. Those from Category B are from a distribution with a mean orientation that is horizontal.

accounts of PC is the generalized context model (GCM) proposed by Nosofsky and colleagues (McKinley & Nosofsky, 1995; Nosofsky, 1984, 1986, 1992, 2000; Zaki, Nosofsky, Stanton, & Cohen, 2003). The GCM assumes that responses to a categorization probe p entail a feature-matching process that begins by determining the probe's similarity to each previously encountered exemplar a of Category A and each previously encountered exemplar b of Category B. Similarity is expressed as an exponential decay function of Euclidean distance d in a multidimensional feature space (these values typically are obtained via multidimensional scaling), weighted by a constant c:

$$s_{pj} = e^{(-cd_{pj})}. \qquad (1)$$

In other words, the psychological similarity between a probe p and a given exemplar j in memory decreases exponentially with increases in physical distance in feature space.

The similarity values for a probe relative to each stimulus in Category A are summed and divided by the summed similarity values between the probe and items from all studied categories K (here, K contains all exemplars from both A and B). This operation, an application of Luce's choice ratio, predicts the categorization probability for a given stimulus:

$$P(A|p) = \Sigma_{a:A} s_{pa} / \left[\Sigma_{k:K} s_{pk} \right]. \qquad (2)$$

Kahana and Sekuler (2002) extended the GCM's feature-matching approach to VSTM. This approach had previously been used in so-called global memory models of long-term memory (Clark & Gronlund, 1996). As the core assumption of this large class of models represents a form of summary statistic, a brief explanation of that assumption may be useful. Early models of recognition memory assumed that the receipt of a test probe stimulus elicited some signal of that

stimulus's familiarity (Green & Swets, 1966; Wickelgren & Norman, 1966). Presumably, the strength of this signal was used to distinguish novel from repeated occurrences of an item. In recognition memory research, novel and repeated occurrences are known as lures and targets, respectively. This simple traditional account of recognition runs headlong into two problems, one conceptual and one empirical. Conceptually, this account could work if only a single item had been presented and held in memory. It is easy to imagine that under this condition, a subject could match the memory representation of that one item to the probe. In fact, this is the kind of single one-to-one comparison that lies at the heart of much psychophysics, including the widely used same–different paradigm (Macmillan & Creelman, 2005). However, this simple account cannot be scaled up to deal with what happens when multiple study items have been presented and consequently multiple items are held in memory. How, from the multiple memories stored on a trial, could one critical memory be isolated and used to generate the requisite familiarity signal?

Empirically, this simple account could also not explain why recognition judgments are governed not only by a probe's similarity to the lone study item to which it was most similar but are also influenced by the probe's similarity to other study items. As Sekuler and Kahana (2007) have noted, global matching models were introduced to address both these challenges. Such models assume that a recognition judgment reflects just a single value, which summarizes or aggregates information about the multiple items that are stored in memory. Some comparison of the probe stimulus to the mean of this summary information could be the basis for the recognition judgment, and possibly also the basis of the familiarity signal posited by other models.

Kahana and Sekuler (2002) adapted the basic GCM framework, including a global memory core, in order to model response rates in a simple Sternberg visual-recognition paradigm. On each trial, subjects were shown a sequence of either one, two, or four visual textures (multicomponent Gabors). This sequence of study items, whose members varied from trial to trial, was followed by a probe texture. Subjects were tasked with judging whether the probe texture had been among the items in the study set. Through careful control of stimulus parameters such as the phase, contrast, and vertical and horizontal spatial frequencies of the textures' luminance profiles, the authors were able to obtain interstimulus distances and similarity measures as in studies of multidimensional PC.

Responses were modeled with what has become known as the noisy exemplar model (NEMO). This model shares the GCM's (and various global matching models') assumption that response rates reflect a pairwise similarity computation between the probe on a

given trial and the representations of multiple items stored in memory. As explained previously, such a computation may allow subjects to discriminate between previously seen target items and new items, as similarity will generally be greatest for probes that match recently encountered items and lowest for probes that mismatch. However, NEMO departs from the GCM in a number of key ways, which track key differences between the mechanisms of VSTM and those of PC. First, NEMO assumes that the contribution of similarity to recognition judgments depends not upon a Luce-ratio computation but upon a linear sum. The similarity between a vector containing probe features and the vector **p** of features for study item s_i is defined by NEMO as

$$\text{Similarity} = \Sigma_{i:L}\alpha_i\eta(\mathbf{p}, \mathbf{s_i} + \varepsilon), \tag{3}$$

where α is a memory-decay weight, η is summed over a total of L items, and ε represents an additional noise factor. The value of η follows a similarity gradient similar to that in the GCM:

$$\eta(\mathbf{p}, \mathbf{s}_i) = e^{-\tau d(\mathbf{p},\mathbf{s}_i)^c}. \tag{4}$$

Another key difference between NEMO and the GCM is NEMO's inclusion of a term representing the similarity of individual study items to one another—i.e., the homogeneity of items in the study set:

$$\text{Homogeneity} = \frac{1}{L(L-1)}\beta\Sigma_{i:L-1}\Sigma_{j:L}\eta(\mathbf{s_i} + \varepsilon, \mathbf{s_j} + \varepsilon). \tag{5}$$

In fits to data from the Sternberg scanning task (Sternberg, 1966), Kahana and Sekuler (2002) consistently found effects of homogeneity, ruling out simpler exemplar models that included only the similarity term. Specifically, they concluded that the best model for their data was one that first summed the similarity and homogeneity terms and then compared the result to a response criterion C_L. In this NEMO model, subjects respond yes when

$$\text{Similarity} + \text{Homogeneity} > C_L. \tag{6}$$

As detailed in Equations 3–6, on each trial NEMO generates a binary response that can be averaged over individual trials, just as individual subjects' responses are. Specifically, each trial is assumed to involve a summed similarity calculation, where the result is compared to the response criterion C_L. Since its inception, several of NEMO's key predictions have been verified, and the model itself has been extended to behavioral and electrophysiological results from various tasks (van Vugt et al., 2009; van Vugt, Sekuler, Wilson, & Kahana, 2013; Yotsumoto, Kahana, Wilson, & Sekuler, 2007).

In effort to extend GCM to VSTM, Nosofsky, Little, Donkin, and Fific (2011) proposed an exemplar-based random-walk model (Nosofsky & Palmeri, 1997) for VSTM. This extension entails the strong claim, which is embodied in the GCM approach, that individual exemplars alone determine responses to recognition probes. The model has no parameter that could accommodate a role for summary statistics such as homogeneity or perceptual averages. However, the assumption that individual exemplars alone determine short-term recognition is clearly at odds with the large body of findings showing that summary statistics do in fact influence VSTM responses. As we have described already, the influence of perceptual averaging and related prototype effects in VSTM is well established. In fact, the literature on perceptual averaging has progressed far beyond simple demonstrations that such averaging influences VSTM; it has begun to converge on the functional significance of such mechanisms and the contexts under which they operate (see contributions elsewhere in this issue). NEMO may not include a perceptual averaging mechanism per se, but it does incorporate the influence of a summary statistic, namely a computation of the homogeneity of all the study items that are to be remembered.

Exemplar-based models often specify a mechanism by which a subject could recognize that a visual aspect of a stimulus matches the visual aspect of a different, recently seen stimulus. However, such models usually include fewer specifics of the mechanisms by which recognition of a stimulus's features improves with repeated presentations of that stimulus—in other words, learning. Recently, Gold, Aizenman, Bond, and Sekuler (2014) devised a way to simultaneously examine short-term memory and longer-term, incidental learning. In their hands, the term *incidental* signified learning that "apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in question" (McGeogh & Irion, 1953).

On each trial, subjects were shown a rapidly presented sequence of eight quasi-random luminances, all delivered to the same central region of a display. On some trials, the final four luminances in a sequence repeated the first four identically (repeat trials), while on other trials, the final four luminances were uncorrelated with the first four (nonrepeat trials). Subjects' task was to categorize a sequence as repeat or nonrepeat. The luminances' rapid presentation rate (8 Hz), coupled with the relatively narrow range from which luminances were sampled, made this task quite challenging. In some experiments, one randomly chosen repeat sequence, constructed independently for each subject, was presented multiple times over the course of a block of trials. Recurring presentations were randomly interspersed among presentations of other stimuli, both repeat and nonrepeat. Subjects were
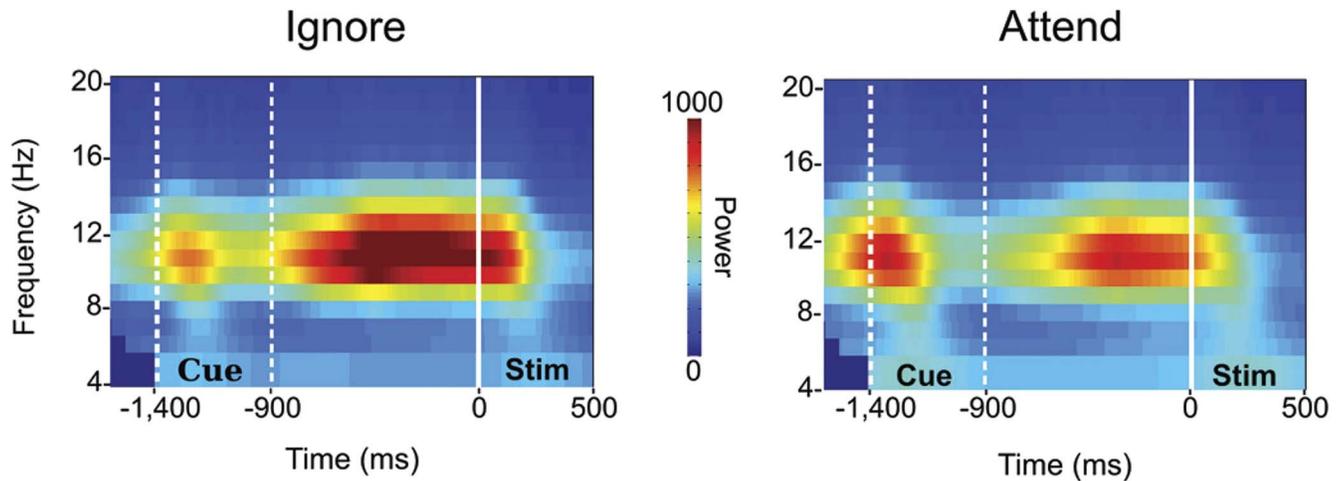
Figure 4. Time-frequency plot showing the effects of a cue to either ignore (left panel) or attend (right panel) an upcoming visual stimulus. The plot shows an increase in oscillatory power in the alpha band (8–14 Hz) of subjects' electroencephalographs immediately following offset of the cue to ignore the stimulus. Reprinted from "Attention-Modulated Alpha-Band Oscillations Protect Against Intrusion of Irrelevant Information" by L. Payne, S. Guillory, and R. Sekuler, 2013, *Journal of Cognitive Neuroscience*, *25*, p. 1469. Copyright 2013 by the Massachusetts Institute of Technology. Adapted with permission.

not told that some sequence would recur, and the relative homogeneity of the stimulus pool from which all sequences were drawn made it hard for them to spot the recurrence of a particular sequence.

Despite not noticing that a sequence recurred intermittently, and despite the fact that spotting a recurrence was actually irrelevant to the short-term memory task (judging whether the final four items in a sequence did or did not repeat the first four), the subjects showed gradual improvement on the short-term memory task over multiple presentations of the same repeat stimulus—that is, incidental learning was observed. An analysis of the luminance sequences presented on each trial revealed that when judging whether a sequence was repeat or nonrepeat, subjects might have made unusual use of summary statistics. Specifically, their performance tracked the difference between the summed luminances of a sequence's first four items and the summed luminances of that sequence's last four items. The researchers conjectured that after computing the two summary statistics for each sequence, subjects would, if the values of those two summary statistics were sufficiently close to one another, deem the sequence to have been repeat. In this case, therefore, the stimulus and task promoted an unusual use of perceptual averaging: Subjects extracted not just a single summary statistic from a stimulus but two, which were then compared to one another. Note that this result is consistent with earlier findings suggesting memory for multiple prototypes in perceptual categorization (Morgan et al., 2000), and supports the simple prototype model of categorization described in the previous section. Specifically, that model assumed that perceptual averages are computed separately for categories A and B and maintained in memory to determine future categorization decisions.

Neural mechanisms of averaging

Although some computational models of perceptual averaging's neural basis have been advanced (e.g., Koene, 2006), data that bear directly on the neural mechanisms of perceptual averaging are lacking. However, some electroencephalographic indices appear to reflect conditions under which perceptual averages are most likely to impact VSTM. Perhaps the most robust and widely studied of these indices is the alpha band (8–14 Hz). A large literature now exists demonstrating that oscillatory power in this band is predictive of stimulus encoding (Romei, Gross, & Thut, 2010) and appears to covary with attentional engagement in both the visual (Snyder & Foxe, 2010; Zanto et al., 2011) and auditory (Banerjee, Snyder, Molholm, & Foxe, 2011; Dubé, Payne, Sekuler, & Rotello, 2013) domains (for a review, see Payne & Sekuler, 2014). A representative result, from a recent study by Payne, Guillory, and Sekuler (2013), is displayed in Figure 4. The figure shows increased alpha power following offset of a cue to ignore an upcoming stimulus (left panel), relative to a condition in which subjects were cued to attend the upcoming stimulus (right panel).

The work by Payne et al. (2013) demonstrates that alpha power predicts not only visual encoding and short-term recognition but also the degree of feature blending between sequentially presented stimuli (Dubé et al., 2014). In that experiment, two Gabor patches were presented sequentially for study, along with attention-
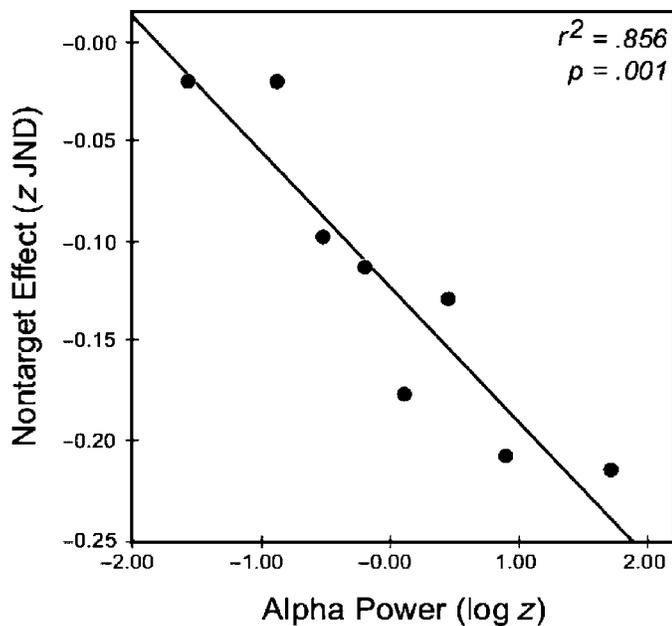
Figure 5. Scatterplot illustrating the negative relationship between poststimulus alpha power and the distorting effect of that just-presented, task-irrelevant stimulus. This association suggests an increased influence of perceptual averages on VSTM responses when attentional control is compromised. Adapted from "Attention-Modulated Alpha-Band Oscillations Protect Against Intrusion of Irrelevant Information," by L. Payne, S. Guillory, and R. Sekuler, 2013, *Journal of Cognitive Neuroscience, 25*, p. 1471, figure 7B. Copyright 2013 by the Massachusetts Institute of Technology. Adapted with permission.

directing cues as in the experiment by Dubé et al. (2014) described previously. However, rather than ask for a recognition response to a probe Gabor, the researchers turned to the method of adjustment: Subjects used a slider on a graphical user interface to adjust a probe Gabor's spatial frequency so that it matched the remembered spatial frequency of the cued study Gabor. Each response provided a measure of recall accuracy. Errors in reproduction were expressed in JND units for each subject and then sign-corrected so that errors of reproduction in the direction of the irrelevant study item's spatial frequency were given a positive sign and errors in the opposite direction were given a negative one. Subjects were cued in advance as to which of the trial's two study items would be relevant.

Payne et al. observed that the visual characteristics of the irrelevant item promoted errors in the reproduction of the relevant, target item in memory. In these errors, the spatial frequency of the recalled visual stimulus shifted toward the spatial frequency of the study item that had been irrelevant. Though not explicitly described as a perceptual averaging effect, this result provides an important glimpse into the process by which perceptual averages may be con-

structed. It may also identify a neural signature for conditions under which perceptual averages are most likely to influence VSTM responses.

Importantly, Payne et al. also found that alpha power immediately following the presentation of the irrelevant study item was negatively correlated with the size of that irrelevant or nontarget item's distorting effect on recall, as indicated in Figure 5. A similar effect was found in the prestimulus interval, consistent with prior work, though this effect was only marginally significant.

In conjunction with the large body of results that show a strong predictive relationship between alpha power and retrieval from both VSTM (Huang & Sekuler, 2010a) and long-term recognition (Dubé et al., 2013), the findings of Payne et al. indicate that posterior alpha power can be used to assess the level of attentional engagement on a given recognition trial. Moreover, posterior alpha power appears to predict the degree of influence exerted by perceptual averages. Such effects are most likely to occur when visual attention is not (or cannot be) focused on the current, task-relevant stimulus. Such a state is reflected in relatively high posterior alpha power in the period around stimulus encoding.

Event-related potentials (ERPs), too, show promise as neural markers of perceptual averaging. For example, a recent study by Zanto, Sekuler, Dube, and Gazzaley (2013) used ERPs to evaluate Ball and Sekuler's neural-biasing hypothesis, which holds that directional expectation can bias brain mechanisms that process motion stimuli. More specifically, Ball and Sekuler (1980) hypothesized that successive encounters with multiple RDCs built up an expectation for subsequent directions of motion, and that this expectation reflected an adaptive change in the tuning of motion-sensitive cortical mechanisms. The result, on this view, was a neural representation that had particular sensitivity to the average of the individual RDC directions that were encountered across trials. To test this neural-biasing hypothesis, Zanto et al. supplemented the RT measurements used by Ball and Sekuler with electroencephalographic measurements, incorporating both into a variant of Ball and Sekuler's paradigm. The design involved presentation of an RDC on each trial, including one direction that was presented on the majority of trials (and that approximated the average of the RDC directions), and a number of alternative (and less frequently presented) RDCs presented on other trials. As in the Ball and Sekuler study, participants were required to make a detection response to the onset of RDC motion on each trial.

The analysis of electroencephalographic responses to motion onset focused on the amplitude of an early (N1) component of ERPs from sensors located over three posterior-occipital regions of interest. This ERP is

sensitive to manipulations of visuospatial attention, with larger N1 amplitudes occurring when stimuli appear at attended (as opposed to unattended) locations (Hillyard & Anllo-Vento, 1998), and is thought to reflect a discriminative process within the focus of attention (Vogel & Luck, 2000). Zanto et al. found that the amplitude of the N1 response to motion onset tracked the angular difference between the RDC's actual direction of motion and the expected direction, with amplitudes decreasing with increases in the difference. The effect was parabolic, mirroring the pattern in the associated RTs as well as those obtained with Ball and Sekuler's paradigm (Figure 1). This result supports the claim that Ball and Sekuler's result reflects the influence of neural biasing toward the average RDC direction, possibly an adaptive response that reduces the effect of stimulus uncertainty.

Conclusions

In this review we have discussed a wide range of evidence suggesting that perceptual averaging is a basic and obligatory aspect of the visual memory system. Such averages appear to be computed even in the absence of a requirement to do so, and have greatest influence when subjects are most likely to be uncertain as to the stimulus conditions or the fidelity of their memory representations. Though little is known about the neural computations that result in perceptual averages, evidence has begun to accumulate suggesting that conditions of divided attention or disengagement of attention produce the greatest reliance on perceptual averages. This makes sense given (a) Ariely's (2001) finding that memory for a perceptual average may be preserved even when memory for individual items is reduced to chance and (b) demonstrations that reliance on perceptual averages compensates for information loss in VSTM and may actually improve performance on VSTM tasks (Alvarez, 2011; Wilken & Ma, 2004).

Despite all these findings, strong claims have recently been made that VSTM relies entirely on member identification of individual items held in memory (Nosofsky et al., 2011). The contrast between such a claim and the idea that VSTM makes use of summary statistics could not be more clear. Future work will be required to evaluate such claims fully, determining whether what appear to be the effects of perceptual averages might in the end reduce to the effects of memory for fine details of individual stimuli.

Keywords: ensemble encoding, perceptual averaging, prototype effect, exemplar models, recognition memory

Footnotes

[1] Whether perceptual averaging in a given task occurs at encoding or later, during storage or retrieval, is an open question. For present purposes, we focus on effects of averages whose entry into storage influences visual short-term memory responses regardless of the manner in which those averages were stored.

[2] As mentioned previously, spatial frequency and relevant/irrelevant status of the study items were counterbalanced so that each item status was equally often of a 4- or 12-JND spatial frequency. Results were the same whether the relevant item was of 4 or 12 JNDs of spatial frequency, so to simplify the visual presentation, the trials on which the relevant item had 12 JNDs of spatial frequency were rescaled to 4 JNDs and the corresponding irrelevant items rescaled to 12.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21,* 560–567.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15,* 122–131.

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

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, USA, 106,* 7345–7350.

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

Azzalini, A. (1986). Further results on a class of

distributions which includes the normal ones. *Statistica, 46,* 199–208.

Ball, K., & Sekuler, R. (1980). Models of stimulus uncertainty in motion perception. *Psychological Review, 87,* 435–469.

Banerjee, S., Snyder, A. C., Molholm, S., & Foxe, J. J. (2011). Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: Supramodal or sensory-specific control mechanisms? *The Journal of Neuroscience, 31*(27), 9923–9932.

Busemeyer, J., & Myung, I. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 3–11.

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

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891–900.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics, 76*(3), 746–758.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138*(2), 289–301.

Corbett, J. E., & Song, J.-H. (2014). Statistical extraction affects visually guided action. *Visual Cognition, 22*(7), 881–895.

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20*(2), 211–231.

Dubé, C., Payne, L., Sekuler, R., & Rotello, C. M. (2013). Paying attention to attention in recognition memory: Insights from models and electrophysiology. *Psychological Science, 24,* 2398–2408.

Dubé, C., Zhou, F., Kahana, M., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research, 96,* 8–16.

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70,* 946–954.

Frisby, J. P., & Stone, J. V. (2010). *Seeing: The computational approach to biological vision* (2nd ed.). Cambridge, MA: MIT Press.

Gold, J. M., Aizenman, A., Bond, S. M., & Sekuler, R. (2014). Memory and incidental learning for visual frozen noise sequences. *Vision Research, 99,* 19–36.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences, USA, 97*(22), 12380–12384.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: John Wiley and Sons.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expressions over time. *Journal of Vision, 9*(11):1, 1–13, doi:10.1167/9.11.1. [PubMed] [Article]

Haberman, J., & Whitney, D. (2012). Ensemble perception: summarizing the scene and broadening the limits of visual processing. In J. M. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York: Oxford University Press.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences, USA, 95*(3), 781–787.

Huang, J., & Sekuler, R. (2010a). Attention protects the fidelity of visual memory: Behavioral and electrophysiological evidence. *The Journal of Neuroscience, 30*(40), 13461–13471.

Huang, J., & Sekuler, R. (2010b). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*(2):24, 1–27, doi:10.1167/10.2.24. [PubMed] [Article]

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*(3), 188–196.

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research, 42*(18), 2177–2192.

Koene, A. (2006). A model for perceptual averaging and stochastic bistable behavior and the role of voluntary control. *Neural Computation, 18,* 3069–3096.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt, Brace, and World.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*(1), 49–70.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience, 16*(4), 493–498.

McGeogh, J. A., & Irion, A. L. (1953). *The psychology of human learning.* New York: Longmans, Green.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 128–148.

Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research, 40*(17), 2341–2349.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–61.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nosofsky, R. M. (2000). Exemplar representation without generalization? Comment on Smith and Minda's (2000) Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1735–1743.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review, 118*(2), 280–315.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*(2), 266–300.

Payne, L., Guillory, S., & Sekuler, R. (2013). Attention-modulated alpha-band oscillations protect against intrusion of irrelevant information. *Journal of Cognitive Neuroscience, 25*(9), 1463–1476.

Payne, L., & Sekuler, R. (2014). The importance of ignoring: Alpha oscillations protect selectivity. *Current Directions in Psychological Science, 23*(3), 171–177.

Pollack, I. (1974). Within- and between-modality correlation detection. *The Journal of the Acoustical Society of America, 55*(3), 641–644.

Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipitoparietal areas in visual input regulation: Correlation or causation? *The Journal of Neuroscience, 30*(25), 8692–8697.

Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science, 16,* 305–310.

Snyder, A. C., & Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: A high-density electrical mapping study. *The Journal of Neuroscience, 30*(11), 4024–4032.

Sternberg, S. (1966, Aug 5). High-speed scanning in human memory. *Science, 153*(736), 652–654.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*(4–8), 411–443.

van Vugt, M. K., Schulze-Bonhage, A., Sekuler, R., Litt, B., Brandt, A., Baltuch, G., & Kahana, M. J. (2009). Intracranial electroencephalography reveals two distinct similarity effects during item recognition. *Brain Research, 1299,* 33–44.

van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2013). An electrophysiological signature of summed similarity in visual working memory. *Journal of Experimental Psychology: General, 142*(2), 412–425.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*(2), 190–203.

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*(5), 931–941.

Watamaniuk, S. N. J., & Sekuler, R. (1992). Temporal and spatial integration in dynamic randomdot stimuli. *Vision Research, 32*(12), 2341–2347.

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research, 29*(1), 47–59.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3*(2), 316–347.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12):11, 1120–1135, doi:10.1167/4.12.11. [PubMed] [Article]

Yotsumoto, Y., Kahana, M. J., Wilson, H. R., & Sekuler, R. (2007). Recognition memory for realistic synthetic faces. *Memory & Cognition, 35*(6), 1233–1244.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1160–1173.

Zanto, T. P., Pan, P., Liu, H., Bollinger, J., Nobre, A., & Gazzaley, A. (2011). Age-related changes in orienting attention in time. *The Journal of Neuroscience, 31,* 12461–12470.

Zanto, T. P., Sekuler, R., Dube, C., & Gazzaley, A. (2013). Age-related changes in expectation based modulation of motion detectability. *PLoS One, 8,* 1–10.