Obligatory and adaptive averaging
in visual short term memory

Chad Dubé*, Robert Sekuler†
*University of South Florida
†Volcan Center for Complex Systems, Brandeis University

Abstract
We evaluate the contention that perceptual averaging is an adaptive and
obligatory mechanism for visual memory. In support of this idea, we review
a wide range of evidence, including findings from our own lab, establishing
that time- and space-averaged memory representations influence detection
and recognition responses, and do so in the absence of any instruction to
compute or report an average. Some of the reviewed work 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
that vary in the degree to which summary statistics impact VSTM, and
discuss a neural signature for contexts in which perceptual averaging exerts
maximal influence.

In its broadest sense, a representation is 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 that is being represented. At one extreme, such representations are
detailed and precise, faithfully capturing a great many of a stimulus’ features; at the other
extreme, they can 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, which forces the system
to fall back on a space- or time-averaged summary of the incoming stimulus. McDermott,

Supported in part by CELEST, an NSF Science of Learning Center (SBE-0354378), and by AFOSR
grant FA9550-10-1-0420. Correspondence should be addressed to chaddube@usf.edu.
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 was presented so quickly that it overwhelmed the ability to encode and recall the information in detail. Finally, a compact representation lends itself to a very important function, communication of what you are seeing. Under most circumstances, it is more helpful 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, unnecessary detail can interfere with communication.

In vision research, analogous effects have been demonstrated in several different arenas. For example, such effects are seen in experiments with random dot cinematograms, which comprise hundreds of spatially intermingled 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 & Sekuler, 1992; Watamaniuk, Sekuler, & Williams, 1989; Watamaniuk & Duchon, 1992).

As Haberman and Whitney (2012) 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, the basic structural mechanisms and functional significance of summary perceptual statistics remain unclear, despite the recent increase in efforts to understand the statistical representations generated by the visual system. In what follows, we summarize some studies that suggest perceptual averaging is an adaptive mechanism, which buffers the visual memory system against interference from old or new information. 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 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 displays comprising multiple circles differing in diameter. 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 2AFC (two probes). Since results with the two tasks were analogous, we will describe the Yes/No variant. Subjects judged the test stimulus in one of two ways. On some trials, subjects made a Yes/No response as to whether the probe was in the memory set (Member Identification). On others, the Yes/No 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, subjects’ mean discrimination performance was quite good. Ariely concluded that the visual system extracts both information about individual items and an average computed over many items. He speculated that when
member information is lost, subjects can fall back onto some summary statistic, like the perceptual average.

Using a variety of tasks, other studies of VSTM have explored different aspects of this putative averaging process. Among the tasks used in those studies were 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 involving brief presentation of visual displays followed by subjects’ estimates of some average feature of the just-presented stimuli (Emmanouil & Treisman, 2008; Chong & Treisman, 2005a, 2005b). 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 in space, as in Ariely’s and so many other studies, but 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). This important finding links perceptual averaging to earlier studies that suggest 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 vectors of items, or from temporally-defined ones. This latter 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 could and does extend to events and items that span such boundaries.

In 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 must explicitly 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 whether averaging is activated and driven by the instruction to compute such an average. Recent work from our laboratory (discussed below), 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

Studies of perceptual averaging dating back more than three decades revealed the contribution of averaging under conditions in which subjects were not asked to compute an average. In one such study, Ball and Sekuler (1980)’s 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 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 showed a U-shaped pattern, increasing
as the directions of the two directions that were possible on any trial diverged from one another.

The authors tested the idea that subjects may monitor visual channels that are maximally responsive to the direction midway between the two alternatives in uncertainty conditions. 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 whose direction matched the perceptual average of the two directions, and which was presented on only 5 to 10% of trials. Such an infrequent stimulus would normally produce inflated RTs. 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 supports Ball and Sekuler’s hypothesis that a representation had been encoded and maintained over trials of the direction that lay midway between the two remembered RDC directions. The authors hypothesized that such representations could be used to mitigate uncertainty about the stimulus that might be presented on a given trial.

Ball and Sekuler’s neural-biasing hypothesis suggests that expectation biases brain mechanisms that process motion stimuli. More specifically, they hypothesized that this expectation effect reflected an adaptive change in the tuning of motion sensitive cortical mechanisms. To test this neural biasing hypothesis, Zanto, Sekuler, Dubé, and Gazzaley (2013) supplemented the measurements of reaction time to motion onset, with electroencephalographic measurements, incorporating both into a variant of Ball and Sekuler’s paradigm. Their analysis of EEG 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. Zanto et al. found that the amplitude of the N1 to motion onset tracked the angular difference between the RDC’s actual direction of motion and the direction that previous stimuli had led the subject to expect. This result directly supports the claim that Ball and Sekuler’s result reflects the influence of neural biasing, possibly an adaptive response that reduces the effect of stimulus uncertainty.

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 puts pressure on the subject to compute such a measure. Would such a statistic be computed without such pressure? Adapting a procedure developed by Huang and Sekuler (2010a), Dubé et al. (2014) recently addressed this in a study of stimulus uncertainty in which no requirement to report an average was included. On each trial, subjects saw two Gabor patches (sinusoidally-modulated, windowed luminance gratings) presented sequentially. Each Gabor comprised both vertically- and horizontally-oriented sinusoid components. Following the second Gabor, a recognition probe was presented that either did or did not match one of the two study Gabor s. Subjects then made a Yes-No recognition response to the probe.

The Gabor stimuli were constructed in such a way that only the vertical spatial frequency dimension of each Gabor pattern could be used to make a response. The precise values of spatial frequency for each subject were obtained using a staircase technique, allowing each stimulus to be expressed in just noticeable difference (JND) units above a fixed
Figure 1. RT tuning curve from Ball and Sekuler (1980) showing changes in simple reaction time responses to detect the onset of upward (90°) motion in conditions where that direction was presented on every trial (90°) or in which one of various alternate directions could also be presented on a trial. Reproduced from Ball and Sekuler (1980)

base frequency. The recognition probes presented to each subject in the experiment took on 15 different JND values, including the value that matched the relevant study item (Target trials) 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), which allowed separate estimates of the Gaussian variance and skew of the resulting response distributions.

A key manipulation in the experiment involved the presentation of an attention-directing cue. This cue was inserted at 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 (Pre-Cue condition), between the two items (Mid-Cue), or after both items had been presented (Post-Cue).

The results are shown in Figure 2, with the target’s position re-scaled to 4 JNDs,
abilities at each probe frequency: 1989 be described as the sum of the model-predicted vectors corresponding to the two
for the two
Fig. 3. 

Panel A: The same conditions, but for trials on which the task-relevant item was the second study item. Reproduced from Dubé, Zhou, Kahana, and Sekuler (2014).

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. Panel A: Three selective attention conditions and a baseline (Single), when the first study item was task-relevant. Panel B: The same conditions, but for trials on which the task-relevant item was the second study item. Reproduced from Dubé, Zhou, Kahana, and Sekuler (2014).

and the irrelevant study item’s position to 12 JNDs, in order to facilitate visual inspection. Included are data from a baseline condition, “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 (scaled to 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 served as 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 to the relevant stimulus in advance (Post-Cue condition), thereby producing maximal uncertainty as to the fidelity of a given memory representation. Dubé et al. concluded that perceptual averaging is a basic, perhaps obligatory operation for the visual memory system.

Prototypes and Exemplars in VSTM

The prevalence of perceptual averaging effects, even in the absence of instruction, has theoretical importance for researchers interested in perceptual categorization (PC). Early studies of categorization have 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). However, the most successful class of models
for PC have arguably been those that assume individual exemplars, rather than prototypes computed over individual exemplars, influence categorization responses (Nosofsky, 1992).

Consider a simple example scenario in which subjects study sets of simple stimuli (e.g., circles) which differ in one feature (e.g., orientation of a line drawn along the circle’s radius). Two classes of stimuli exist, unbeknownst to the subject: One class, “Category A”, has a distribution of orientations centered at 90°, and the other, “Category B”, has a distribution centered at 180°. The subject must make a classification response (“A” or “B”) for each stimulus, and is provided feedback after each trial, allowing them to learn the category assignments over several trials. One way of modeling subjects’ performance of 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 have argued strongly against prototype models, and in favor of exemplar models of PC (see Murphy, 2002 for review). Perhaps the most influential of these accounts of PC is the Generalized Context Model (GCM) proposed by Nosofsky and colleagues (McKinley & Nosofsky, 1995; Nosofsky, 1992, 1984, 1986, 2000; Zaki, Nosofsky, Stanton, & Cohen, 2003). The GCM assumes that responses to a categorization probe \( p \) entail a feature matching process that compares the probe’s features to those of each previously-encountered exemplar \( a \) of category A and \( b \) 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, \( c \):

\[
s_{pj} = e^{-cd_{pj}}
\]  

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 application of Luce’s Choice Ratio predicts the categorization probability for a given stimulus. 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 A and B).

\[
P(A|p) = \frac{\Sigma_{a:A}s_{pa}}{[\Sigma_{k:K}s_{pk}]}
\]  

Kahana and Sekuler (2002) extended the feature-matching approach to VTSM. 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 warranted. Early models of recognition memory assumed that the receipt of a test probe stimulus elicited some signal of that stimulus 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. 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. 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 be 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 was also influenced by the probe’s similarity to other study items. As Sekuler and Kahana (2007) note, 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 is the basis for the recognition judgment, and possibly the familiarity signal that is posited by other models.

Kahana and Sekuler 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 visual textures (multi-component Gabors), either one, two, or four such textures. 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 inter-stimulus 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 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 items stored in memory. As explained above, such a computation may allow subjects to discriminate between previously-seen target items and new items, as similarity will generally be greatest for those probes that match items that have been recently encountered, and lowest for those that mismatch. However, NEMO departs from 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 does not depend upon a Luce ratio computation, but upon a linear sum. This result is the average of similarity values, multiplied by a constant (the memory set size). The similarity between a vector containing probe features, \( \mathbf{p} \), and the vector of features for study item \( \mathbf{s}_i \), is defined by NEMO as:

\[
\text{Similarity} = \sum_{i=1}^{L} \alpha_i \eta(\mathbf{p}, \mathbf{s}_i + \epsilon) \quad (3)
\]

where \( \alpha \) is a memory decay weight, \( \eta \) is summed over a total of \( L \) items, and \( \epsilon \) represents an additional noise factor. \( \eta \) follows a similarity gradient similar to that of the GCM.

\[
\eta(\mathbf{p}, \mathbf{s}_i) = e^{-\tau d(\mathbf{p}, \mathbf{s}_i)^c} \quad (4)
\]
Another key difference between NEMO and GCM is NEMO’s inclusion of a term representing the similarity of individual study items to one another, \textit{i.e.}, the homogeneity of items in the study set:

\[
\text{Homogeneity} = \frac{1}{L(L-1)} \beta \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} \eta(s_i+, s_j+)
\]  \hspace{1cm} (5)

In fits to data from the Sternberg task, Kahana and Sekuler 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:

\[
\text{Similarity} + \text{Homogeneity} > C_L
\]  \hspace{1cm} (6)

As detailed in the equations above, on each trial, NEMO generates a binary response that can be averaged over individual trials, just as individual subjects’ response 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 encompass results, behavioral and electrophysiological, from various tasks (Yotsumoto, Kahana, Wilson, & Sekuler, 2007; van Vugt et al., 2009; van Vugt, Sekuler, Wilson, & Kahana, 2013).

In effort to extend GCM to visual short term memory, Nosofsky, Little, Donkin, and Fific (2011) proposed an exemplar-based random walk model (EBRW, 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. There is no parameter in the model that describes a role for summary statistics such as homogeneity or perceptual averages. However, such strong claims are clearly at odds with the large body of findings showing that summary statistics do in fact influence VSTM responses. As we have described above, 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 of how such averaging influences VSTM, and has begun to uncover the functional significance of such mechanisms and the contexts under which they operate (see contributions elsewhere in this Special Issue). NEMO may not include a perceptual averaging mechanism \textit{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. As for the summed similarity statistic, the homogeneity statistic is the average of similarity values weighted by a constant (the memory set size).

Exemplar-based models often specify a mechanism by which a subject could recognize that a visual aspect of a stimulus matched the visual aspect of a different, recently seen stimulus. However, such models usually include less specifics of the mechanisms by which recognition of a stimulus’ features improves with repeated presentations of that stimulus, in other words, learning. Recently, Gold, Aizenman, Bond, and Sekuler (2014) devised a way to examine simultaneously 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, Gold et al.’s subjects were shown a rapidly presented sequence of eight quasi-random luminances. 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 (Non-Repeat trials). The subjects’s task was to categorize a sequence as Repeat or Non-Repeat. The luminances’ rapid presentation rate (8 Hz) coupled with the relatively narrow range of luminances that could be 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 Non-Repeat. Subjects were 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 subjects 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), Gold et al.’s 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 a Repeat or Non-Repeat, subjects might have made unusual use of summary statistics. Specifically, subjects’ 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. Gold et al. conjectured that after computing the two summary statistics for each sequence, if the values of those two summary statistics were sufficiently close to one another, subjects would deem the sequence to have been a 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 the simple Prototype model of categorization described in the previous section. Specifically, that model assumed that perceptual averages are computed separately and maintained in parallel for different stimulus categories.

Neural Mechanisms of Averaging

Computational models of perceptual averaging’s neural basis have been advanced (Koene, 2006), but direct data on the neural mechanisms of perceptual averaging are lacking. However, electroencephalographic (EEG) indices have been uncovered that appear to reflect those conditions under which averages are most likely to impact VSTM responses. Perhaps the most robust and widely studied of these indices is the (8-14Hz) alpha band. 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 (See Payne & Sekuler, 2014 for review).

A recent study by Payne, Guillory, and Sekuler (2013) demonstrates that alpha power not only predicts visual encoding and short term recognition, but also the degree of feature blending between sequentially presented stimuli (Dubé et al., 2014). In their study, two
Gabor patches were presented sequentially for study, along with attention cues as in the experiment by Dubé et al. described above. However, rather than ask for a recognition response to a probe Gabor, Payne et al. turned to the method of adjustment: subjects adjusted a probe Gabor’s spatial frequency via a slider until the frequency matched the remembered spatial frequency of the cued study Gabor. The result was taken as a measure of recall. Errors in reproduction were expressed in JND units for each subject, and were sign-corrected so that errors of reproduction in the direction of the irrelevant study item’s spatial frequency were positive and errors in the opposite direction were negative. Subjects were cued in advance as to which of the two upcoming study items would be task-relevant on a particular trial.

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 task-irrelevant. Though not explicitly described as a “perceptual averaging” effect, these results nonetheless provide an important glimpse into the process by which blended representations are constructed. It also identifies a neural signature for conditions under which such blended representations are most likely to influence 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 4. A similar effect was found in the pre-stimulus interval, consistent with prior work, though this effect was only marginally significant.

This findings of Payne et al., coupled with the large body of other results showing a strong predictive relationship between alpha power and retrieval from both visual short term memory (Huang & Sekuler, 2010b) and long term recognition (Dubé et al., 2013),
show 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 that is exerted by perceptual averages. Such effects are 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 levels of alpha power surrounding an encoding event, as measured via posterior electrodes.

Figure 4. Scatterplot illustrating the negative relationship between post-stimulus 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 Payne, Guillory, and Sekuler (2013), Figure 7B.
Conclusions

In this review, we have discussed a wide range of evidence that suggests 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 an influence on responding that is greatest under conditions where subjects are 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 visual attention produce the greatest reliance on perceptual averages. This makes sense, given i) the initial findings of Ariely (2001) that perceptual averages are retained even when memory for individual items is reduced to chance and ii) the 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.

References


