Vision Research 96 (2014) 8–16

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Similarity-based distortion of visual short-term memory is due to perceptual averaging

Chad Dubé [a,*,1], Feng Zhou [b,1], Michael J. Kahana [c], Robert Sekuler [d]

[a] University of South Florida, 4202 East Fowler Ave., Tampa, FL 33620, USA
[b] Weaver, Austin, Velleneuve, and Sampson, LLP, 555 12th St., Oakland, CA 94607, USA
[c] University of Pennsylvania, 3401 Walnut St., Philadelphia, PA 19104, USA
[d] Brandeis University, 415 South Street, Waltham, MA 02454, USA

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form 15 December 2013
Available online 4 January 2014

Keywords:
Visual short-term memory
Perceptual averaging
Recognition memory
Distributional analysis
Selective attention

ABSTRACT

A task-irrelevant stimulus can distort recall from visual short-term memory (VSTM). Specifically, reproduction of a task-relevant memory item is biased in the direction of the irrelevant memory item (Huang & Sekuler, 2010a). The present study addresses the hypothesis that such effects reflect the influence of neural averaging under conditions of uncertainty about the contents of VSTM (Alvarez, 2011; Ball & Sekuler, 1980). We manipulated subjects' attention to relevant and irrelevant study items whose similarity relationships were held constant, while varying how similar the study items were to a subsequent recognition probe. On each trial, subjects were shown one or two Gabor patches, followed by the probe; their task was to indicate whether the probe matched one of the study items. A brief cue told subjects which Gabor, first or second, would serve as that trial's target item. Critically, this cue appeared either before, between, or after the study items. A distributional analysis of the resulting mnemometric functions showed an inflation in probability density in the region spanning the spatial frequency of the average of the two memory items. This effect, due to an elevation in false alarms to probes matching the perceptual average, was diminished when cues were presented before both study items. These results suggest that (a) perceptual averages are computed obligatorily and (b) perceptual averages are relied upon to a greater extent when item representations are weakened. Implications of these results for theories of VSTM are discussed.

1. Introduction

Everyday experience provides undeniable proof that memory is fragile and fallible. Importantly, such errors of memory have provided many key insights into memory's structure and properties (Schacter, 2001). One class of memory errors seems especially difficult to explain: errors that suggest memory is retained, but has somehow become distorted. The most-studied of such distortions are ones that afflict long-term memory, but short-term memory is certainly not immune, and may be vulnerable to distorting influences of its own. For example, Huang and Sekuler (2010a) identified a pair of influences that each distorted recall from visual short-term memory. One influence was attributable to the aggregation of stimulus information across trials; the second influence reflected the similarity of a trial's task-irrelevant stimulus to an accompanying task-relevant stimulus. Huang and Sekuler (2010a) described these influences as "attractors" as each seemed to shift or pull subjects' judgments toward itself. Although the timescale of the across-trial effect suggests that it is memory based, Huang and Sekuler (2010a) proposed a different approach for the within-trial effect. They hypothesized that the task-irrelevant stimulus influenced an early, encoding-related process, rather than a later, storage or retrieval process. Specifically, Huang and Sekuler (2010a) posited that subjects stored a weighted average of the *i*th trial's relevant and irrelevant memory items, essentially encoding an ensemble representation for that trial which subsequently influenced recall of the relevant item.

Support for this explanation can be found in studies of perceptual averaging (Alvarez, 2011). Specifically, abundant evidence shows that subjects can accurately compute the average feature values of a set of spatially-separated visual objects. In a seminal study of such averaging, Ariely (2001) showed subjects a display of circles differing in diameter, followed by a test set containing a probe circle. Subjects either indicated whether a probe circle had been in the memory set, or, on some trials, judged whether a test circle was larger or smaller than the mean circle in the memory set. Ariely found that, though memory for individual items was at chance, subjects made accurate mean discrimination responses.

* Corresponding author.
E-mail address: chaddube@usf.edu (C. Dubé).

[1] The first two authors contributed equally to this manuscript.

0042-6989/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2013.12.016

He suggested that the visual system relies on two mechanisms, item identification and averaging, and may discard fine-grained information in favor of the average when it is efficient to do so.

More recent work on perceptual averaging suggests such averages are preserved in conditions of divided or diffuse attention (Albrecht, Scholl, & Chun, 2012; Alvarez & Oliva, 2008, 2009; Chong & Treisman, 2005; but see Emmanouil & Treisman, 2008). Additionally, it has been shown that averages are computed not just over items presented in a particular visual display, but may also be computed over items presented on successive trials (Albrecht & Scholl, 2010; Haberman, Harp, & Whitney, 2009). The latter finding suggests a link between perceptual averaging and the effect reported by Huang and Sekuler, which involved stimuli presented in sequence.

Despite the growing literature on perceptual averaging in VSTM, the conditions determining the influence of such averaging have not been widely explored. What, if anything, determines the extent to which averages are relied upon in VSTM? Assuming that (i) memory load weakens individual VSTM representations and (ii) averages are nonetheless preserved in such conditions, it seems likely that the influence of perceptual averaging may depend (at least in part) on memory load (see, e.g., Ball & Sekuler, 1980, for supporting evidence related to this point). However this is by no means guaranteed: most studies of perceptual averaging explicitly required estimates of such averages. It is not clear that averages are computed obligatorily, let alone that they influence responses that do not nominally require reliance on averages.

The current study addresses two key hypotheses: first, that perceptual averages influence recognition responses in the absence of an explicit requirement to compute averages, and second, that reliance on averages will increase under conditions that promote uncertainty about VSTM's fidelity (Ball & Sekuler, 1980). On most trials of the current experiment, two study items (Gabors) were presented successively, followed after a short delay by a test probe (another Gabor, henceforth denoted $\mathbf{p}$). A visual cue presented at one of several possible times relative to the study items indicated which item, either $\mathbf{s}_1$ (the first stimulus) or $\mathbf{s}_2$ (the second), had to be remembered, and was therefore task-relevant. The subject judged whether the probe's spatial frequency matched the spatial frequency of the task-relevant study item. In order to manipulate selective attention's influence on encoding processes, an attention-directing cue was presented at different times relative to the two study stimuli.

An important aspect of the current design is its inclusion of a graded manipulation of $\mathbf{p} - \mathbf{s}$ similarity. Specifically, we fixed the similarity of $\mathbf{s}_1$ to $\mathbf{s}_2$, and applied a "roving probe" technique to the test probes, in which $\mathbf{p}$ can take on one of several (in this case, 15) similarity values relative to the study items. By using this technique, it is possible to map out *mnemometric functions,* indices of memory strength's distributions (Sekuler & Kahana, 2007; Williams, Titchener, & Boring, 1918; Zhou, Kahana, & Sekuler, 2004). To these distributions, we fit Skew-Normal functions (Azzalini, 1985, 1986; Bansal, Maadooliat, & Wang, 2008), which allowed us to directly measure changes in the spread and skew of the observed mnemometric functions. As we will demonstrate, such distributional analyses provide a powerful window onto the effects of interest, which could be obscured or distorted in an analysis restricted to coarser measures such as mean response rate (see Balota & Yap, 2011, for further arguments in favor of distributional analyses).

To preview the outcome, our distributional analysis of subjects' recognition responses suggests that, under conditions of memory load, subjects rely more heavily on ensemble representations, matching probes to a perceptual average of $\mathbf{s}_1$ and $\mathbf{s}_2$. Thus, our results support a perceptual averaging account of Huang and Sekuler's distortion effect. Furthermore, our results suggest that averages are computed obligatorily, and are relied on in order to compensate for imperfections in VSTM representations.

2. Methods

On each trial two study items, Gabor patches ($\mathbf{s}_1$ and $\mathbf{s}_2$), were presented sequentially in the center of the computer screen, followed by a centrally-presented probe Gabor, $\mathbf{p}$.[2] The Gabor patches varied in spatial frequency, and on each trial a subject made a recognition judgment, reporting whether $\mathbf{p}$'s spatial frequency matched or did not match the remembered spatial frequency of one of the study items. Cues introduced at various times during a trial instructed the subject to attend to one or both of the study items, and to base their recognition judgments on the attended item(s). The cue to direct attention was meant to promote a temporary, top-down attentional selection of a study item that occupied a particular ordinal position in the sequence of study items, thereby strengthening its representation.

The propensity of any lure stimulus (i.e. a non-matching $\mathbf{p}$), to draw false recognitions varies with that lure's similarity to study items: other things being equal, a lure that is similar to one or more study items will attract more false recognitions than one that is dissimilar to the study items (Sekuler & Kahana, 2007). Various experimental manipulations, including manipulations of attention, can alter the proportion and distribution of false recognition responses. Therefore, the set of possible lures in our experiment was constructed so as to maximize sensitivity to the effects in which we were interested. Specifically, we tested subjects with probes whose spatial frequencies roved along the frequency dimension within which $\mathbf{s}_1$ and $\mathbf{s}_2$ were situated.

2.1. Subjects

Six male and eight female paid volunteers completed the experiment; two additional subjects began but failed to complete the multi-session experimental protocol. All subjects had normal or corrected-to-normal Snellen acuity, and normal contrast sensitivity as measured with Pelli–Robson charts (Pelli, Robson, & Wilkins, 1988).

2.2. Apparatus

Stimuli were generated and displayed using Matlab and extensions from the Psychophysics and Video Toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 14-in. cathode ray tube monitor with a refresh rate of 95 Hz, and a screen resolution of 800×600 pixels. Routines from the Video Toolbox were used to calibrate and linearize the display. The mean luminance of the display was maintained at $36\,\mathrm{cd/m^2}$ throughout the experiment.

2.3. Conditions

The experiment comprised eight main conditions in which we varied the stimulus or stimuli that had to be attended and remembered. In all conditions save a single baseline condition described below, on each trial three Gabor patches were presented sequentially ($\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{p}$), for 500 ms each. A three-second retention interval separated the disappearance of $\mathbf{s}_2$ and the onset of $\mathbf{p}$. At the end of a trial, subjects judged whether $\mathbf{p}$ had been among the study items or not. To minimize the possibility that subjects could base judgments exclusively on retinotopic matches between local

[2] A Gabor stimulus comprises a sinusoidally modulated luminance grating windowed by a Gaussian.

features, the absolute phases of both horizontal and vertical components were randomly varied from stimulus to stimulus within a trial. Stimuli on each trial shared a single, common horizontal spatial frequency. As a result, only the other, vertical spatial frequencies' differences were relevant to the task.

Fig. 1 represents the experiment's eight conditions (white ovals) in schematic form. Six of these conditions involved a manipulation of selective attention; two of them did not. We begin with a description of the two baseline conditions, in which attention was not directed to any particular, designated stimulus. In one of these non-selective conditions, which we designate as *Both*, the letter B was presented for 1000 ms before $\mathbf{s}_1$, instructing the subject to attend to and remember both $\mathbf{s}_1$ and $\mathbf{s}_2$, and to judge whether $\mathbf{p}$ matched either study item. In the other non-selective condition, the letter S appeared for 1000 ms at the beginning of a trial, signaling that only a single study item would be presented on that trial. This condition is designated as *Single*. Note that *Single* comprises a simple delayed discrimination task, which provided a baseline against which memory in other conditions could be compared. We expected that the two non-selective conditions, *Single* and *Both*, would produce performance at the upper and lower bounds of recognition, respectively, thereby bracketing the performance generated in the six conditions of selective attention.

In the selective attention conditions, during the 1000 ms interval immediately before the presentation of $\mathbf{s}_1$, between $\mathbf{s}_1$ and $\mathbf{s}_2$, or after $\mathbf{s}_2$, a visual cue, the numeral "1" or "2," was presented for 1000 ms at the center of the display. This numeral cued the subject to attend to either $\mathbf{s}_1$ or $\mathbf{s}_2$, for "1" or "2," respectively. Subjects were instructed that only the attended stimulus item was relevant to their recognition judgment of whether $\mathbf{p}$ did or did not match the cued study stimulus. Six different conditions of selective attention were produced by factorially combining three different times at which a cue could occur and two different study items that could be attended. We designate conditions in which a cue occurred before/between/after the study items as Pre/Mid/Post, respectively. Further, we designate conditions as 1 and 2 according to serial position, first or second, of the attended study item. This system for designating conditions produces six unique binomial designations: *Pre-1*, *Mid-1*, *Post-1*, *Pre-2*, *Mid-2*, and *Post-2*.

A fixation cross remained visible at the center of the screen throughout the experiment except when a stimulus or a cue was being displayed. The conditions of selective attention were tested in randomly-interspersed blocks of 80 trials each. Conditions *Single* and *Both* were run in dedicated blocks of their own; but *Pre-1* and *Pre-2* trials were intermingled in a block, as were *Mid-1* and *Mid-2* in other blocks, and *Post-1* and *Post-2* in other blocks still. Conditions were presented in blocks so as to avoid forcing subjects to make random, trialwise changes in attentional strategy. Importantly, the blocking used here gave subjects no indication as to what item would be task-relevant on a given trial (with the exception of *Single*). Thirteen of the subjects produced a total of 3390 trials each, distributed over 15 sessions; one subject served in additional sessions, producing a total of 6780 trials. This subject's data were consistent with those of the other subjects.

2.4. Stimuli

In each stimulus one vertical and one horizontal sinusoidal-luminance grating were superimposed. As explained below, the spatial frequency of only the vertical component was actually task-relevant. Each sinusoidal component's Michelson contrast was set to 0.2, a value well above the threshold for detection. Each stimulus subtended 6° visual angle at a viewing distance of 82 cm. To minimize edge effects, the luminance gratings were windowed

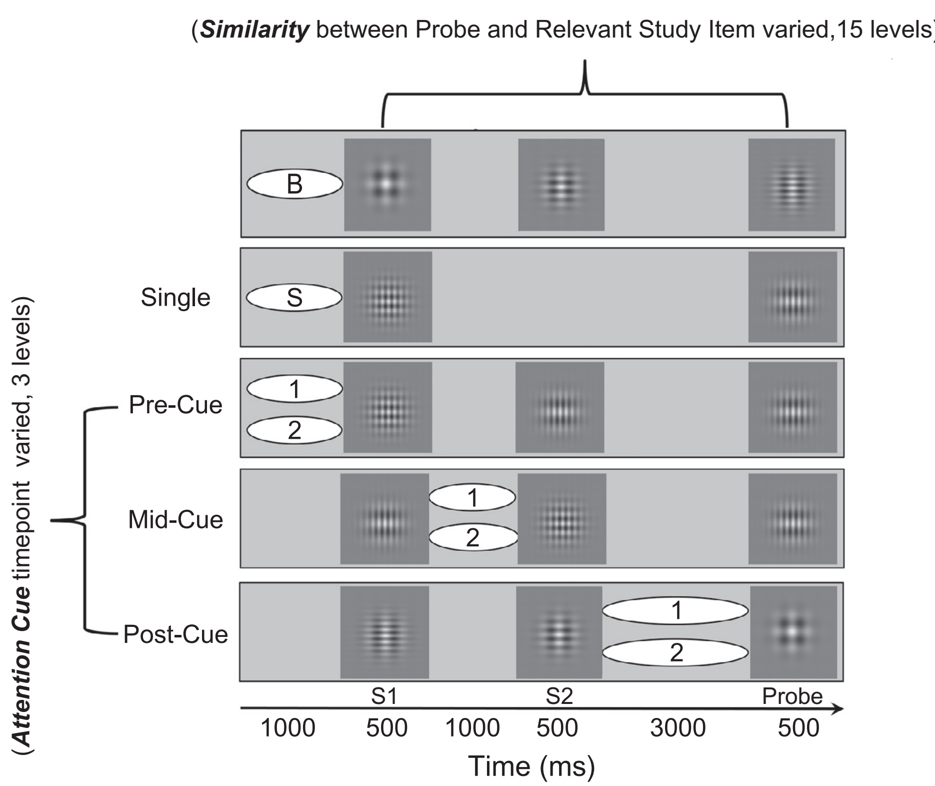

Fig. 1. Schematic representation of sample trials associated with various conditions of stimulus presentation and attentional cues in the experiment. $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{p}$ were presented sequentially. On some trials, the letter "B" appeared before $\mathbf{s}_1$, and the subjects were required to attend to and decide on both $\mathbf{s}_1$ and $\mathbf{s}_2$. On other trials, when the letter "S" appeared at the start of a trial, only a single stimulus $\mathbf{s}$ followed. Finally, in selective attention trials, a numeral, either "1" or "2," occupied one of three intervals, prompting the subjects to attend to $\mathbf{s}_1$ or $\mathbf{s}_2$, respectively.

by a circular 2-D Gaussian function whose space constant was 1° visual angle. Viewing was binocular, with the subject's head supported by a chin-and-forehead rest 114 cm from the display. Additional details of stimulus construction can be found elsewhere (Kahana et al., 2007).

Prior to memory testing, each subject's 2-interval forced-choice spatial frequency discrimination threshold was measured with QUEST, an adaptive psychophysical algorithm (King-Smith et al., 1994; Watson & Pelli, 1983). As implemented here, QUEST identified the difference in spatial frequency that sufficed to produce 75% correct identification of the higher spatial frequency of two gratings. This stimulus normalization was meant to control individual differences in visual encoding as a source of variation in recognition performance (Kahana et al., 2007; Zhou, Kahana, & Sekuler, 2004). The resulting Weber fractions for different subjects ranged from 0.10 to 0.23. Each subject's Weber fraction was then used to normalize the stimuli with which that subject's recognition memory would be tested. Each subject's stimulus set for the recognition experiment comprised Gabors whose sinusoidal components' spatial frequencies were defined by the relation

$$f = f_0(1 + K_s)^n \tag{1}$$

where f_0 is a fixed base frequency, K_s is a subject's Weber fraction, and n is the desired distance from the base frequency, in just-noticeable difference (*JND*) units.

$$K_s = \frac{\Delta f}{f} \tag{2}$$

where Δf is the JND for spatial frequency. As a result, the spatial frequency of evey stimulus could be expressed in JND units.

Within a trial, the horizontal frequency for all three gratings was held constant, but varied between trials, taking random values drawn from a uniform distribution bounded by 0 and 7.0 c/deg. The vertical frequency of each grating was scaled to an individual subject's Weber fraction for spatial frequency. The geometric mean of each trial's study gratings was set to seven *JND*s above a randomly-chosen minimum reference spatial frequency, which ranged from 1.0 to 2.0 c/deg. The value of this geometric mean, in c/deg, varied from trial to trial, so in order to aggregate data over trials, we normalized each trial's geometric mean frequency to the same constant value. The difference between the study gratings' vertical frequencies, $|\mathbf{s}_1 - \mathbf{s}_2|$, was fixed at 8 *JND*s. This substantial difference between study items was meant to minimize interactions between their perceptual representations (Graham, 1989).

Target (**T**) trials, on which **p** matched one of the study items, and *Lure* (**L**) trials, on which **p** matched no study item, occurred with equal frequency. To encourage subjects to attend to and remember both study items, on **T** trials, **p** equally often matched $\mathbf{s}_1$ or $\mathbf{s}_2$. On **L** trials, **p**'s spatial frequency was drawn from a discrete uniform distribution ranging from (i) three *JND*s below the spatial frequency of that trial's lower spatial frequency study item, to (ii) three *JND*s units above the spatial frequency of that trial's higher spatial frequency study item. On half the trials, $\mathbf{s}_1$'s vertical spatial frequency was less than $\mathbf{s}_2$'s; on the remaining trials, the reverse was true. For any trial, **p**'s vertical frequency was chosen from a set of 15 discrete values evenly spaced along a logarithmic scale on which neighboring values differed by one *JND*. Because each trial's minimum possible spatial frequency came from a randomly-chosen value, the actual values of spatial frequency (in c/deg) that could be presented varied from one trial to another. For each trial, **p**'s spatial frequency was drawn from a discrete distribution centered around the geometric mean of that trial's spatial frequencies for $\mathbf{s}_1$ and $\mathbf{s}_2$. On **L** trials each of the 15 possible **p** values was tested 10 times in each session; on **T** trials, each **p** value was tested 75 times.

2.5. Procedure

Subjects initiated each trial by pressing a key on a keyboard linked to the computer that controlled the experiment. Then, after the trial's stimuli had been presented, subjects pressed computer keys corresponding to "Yes" and "No", signaling their judgment that **p** matched or did not match the target stimulus. Subjects had been instructed to respond as accurately and quickly as possible. The computer produced distinctly different tones after correct and incorrect responses, providing trial-wise knowledge of results.

In order to ensure compliance with the instructions to maintain fixation, video images of subjects' eyes were recorded for online monitoring and for subsequent offline analysis. A light-emitting diode not visible to the subject tagged the attentional conditions and the duration of each stimulus. When a live video showed signs of non-compliance, the subject was reminded of the experiment's requirements. One subject had to be excluded from the experiment because of repeated failures to comply with instructions to maintain fixation.

After the experiment was complete, three judges who were blind to the experimental aims viewed sample videos, and identified any shifts in gaze away from the screen. Calibration tests showed that intentional shifts of 1° visual angle were easily detected in the videos. Such shifts were rare, occurring on fewer than 2% of the trials in the video clips examined. Equally important, the frequency of shifts was no higher when an irrelevant stimulus was being presented ($\bar{x} = .015$, *SEM* = .003) than when the stimulus was relevant ($\bar{x} = .018$, *SEM* = .003). We conclude that any effects that might result from shifts of gaze are minimal and not confounding.

3. Results

Subjects' binary recognition responses were transformed into a graphical format that Zhou, Kahana, and Sekuler (2004) termed a mnemometric function. Each mnemometric function expresses the proportion of 'Yes' responses, hereafter P (*Yes*), as a function of **p**'s spatial frequency. Here, **p**'s spatial frequency is defined relative to the spatial frequencies of the study item(s). A mnemometric function reflects the strength of the remembered item or items at the spatial frequency sampled by **p** (see Sekuler & Kahana, 2007; Williams, Titchener, & Boring, 1918; Zhou, Kahana, & Sekuler, 2004).

3.1. Baseline conditions: Single versus Both

We begin with baseline comparisons involving mnemometric functions produced in the two conditions that did not explicitly invoke selective attention. These two conditions were *Single*, with only a single study item presented on each trial, and *Both*, with subjects instructed to give equal attention to two study items. Fig. 2A shows the mnemometric functions for these conditions, with response rates normalized to sum to one in each condition. This was done in order to facilitate the modeling of subjects' functions, and to provide a clearer picture of the distribution of subjects' response rates. Note that as only one study item was presented in *Single*, the designations $\mathbf{s}_1$ and $\mathbf{s}_2$ do not apply to that condition, although spatial frequency (4 or 12 *JND*s) does. Data from the *Both* condition are averaged over the Lower/Higher variable. This designation indicates whether $\mathbf{s}_1$ was lower or higher in spatial frequency than $\mathbf{s}_2$, as there were two items presented in this condition. We averaged over data from conditions of the same spatial frequency, to preserve any differences in spread as a function of Lower/Higher.

Notice first that, as expected, the peaks of each mnemometric function are well aligned with the study items' spatial frequencies.

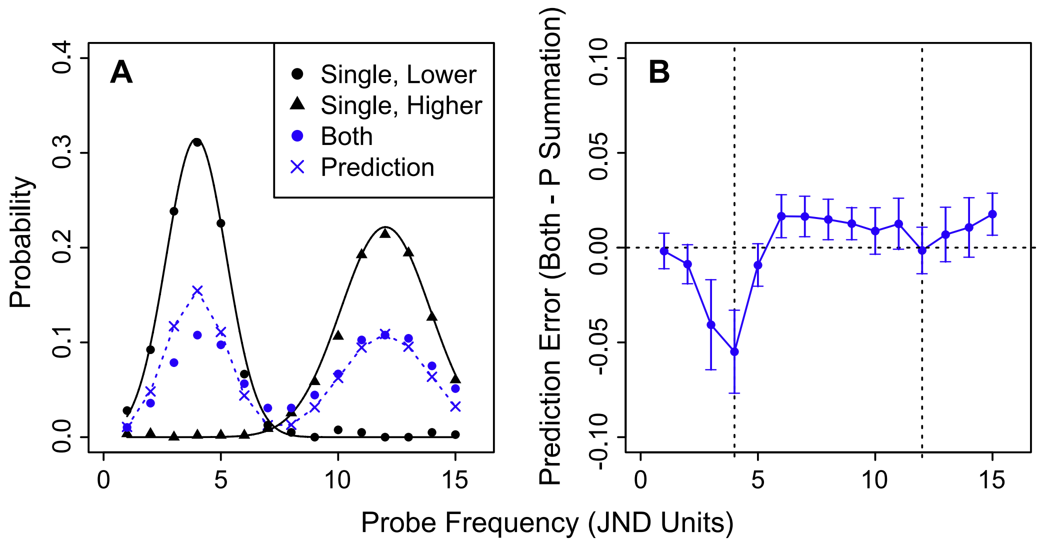

Fig. 2. Mnemometric functions for the two conditions that involved no manipulation of selective attention, *Single* and *Both*, and predictions from the probability summation model. Here, the data from the *Both* condition are averaged over the order of presentation of the Lower and Higher spatial frequency study items (order had little effect on the data). *Panel A*: mnemometric functions for the baseline conditions and probability summation. *Panel B*: predictions for the *Both* condition, derived from probability summation of the two *Single* conditions, expressed as deviations from the observed *Both* data (Both – Probability Summation). The figure shows an inflation in probability density in the region between the values of the study items (vertical dotted lines) of the two *Single* conditions. Error bars represent 95% CIs of the difference at each *JND* point.

In other words, hit rates are highest when **p** actually matches an item in memory. Moreover, peak recognition performance with *Single* ($\bar{x} = 0.79$) is significantly higher than the performance achieved with either study item, $\mathbf{s}_1$ ($\bar{x} = 0.59$) or $\mathbf{s}_2$ ($\bar{x} = 0.67$), in the *Both* condition,[3] $t(13) = 6.48$, $p < .0001$, and $t(13) = 4.98$, $p < .001$, respectively. This advantage shown by the *Single* condition reflects the impact of memory load, that is, the impact of having to hold two items in short-term memory rather than just one. Most importantly, it indicates that the increased memory load did indeed produce weakened memory representations of each study item.

Next, we examined the data for a serial order effect, which is a hallmark of experiments using paradigms like ours. We compared P (*Yes*) for the first and second study items of the *Both* condition. The results show that P (*Yes*) for $\mathbf{s}_2$ was higher than P (*Yes*) for $\mathbf{s}_1$, $t(13) = 2.71$, $p < .05$. This modest but statistically significant recency effect aligns with previous findings produced with similar stimuli and comparable numbers of study items (Kahana & Sekuler, 2002; Kahana et al., 2007; Zhou, Kahana, & Sekuler, 2004).[4]

Turning to the shapes of the distributions, it is clear from inspection of Fig. 2A that false alarm rates in the *Both* condition are elevated for probes that fall near the average of $\mathbf{s}_1$ and $\mathbf{s}_2$ in *JND* units (i.e. near 8 *JNDs*), relative to the corresponding response rates for each of the *Single* conditions. This suggests an influence of perceptual averaging in the *Both* condition, which would produce a VSTM representation whose spatial frequency falls between those of $\mathbf{s}_1$ and $\mathbf{s}_2$.

To assess this potentially important effect, we began by estimating distributional parameters of the *Single* data. To do so, we fit a truncated Skew-Normal distribution (Azzalini, 1985, 1986; Bansal, Maadooliat, & Wang, 2008) to each observed mnemometric function, for each subject. (Truncated distributions were used as responses were only collected over a range of 0–15 *JNDs*). The Skew-Normal distribution is the product of a Gaussian probability density function and its cumulative distribution function. More formally, the Skew-Normal function is defined as:

$$f(x) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right) \quad (3)$$

where ϕ denotes the Gaussian pdf, Φ is the corresponding cdf, and λ is a weight that determines the direction and degree of skewness. Values of $\lambda > 0$ produce positive skew, values < 0 produce negative skew, and $\lambda = 0$ reduces to the normal distribution (in this case, truncated between 0 and 15 jnds). We fit separate truncated Skew-Normals to each *Single* condition, centered on the target *JND* value for each, for each subject. This allowed us to more clearly assess differences between the shapes of the distributions in this and later mnemometric analyses. Fits were obtained by using the *optim* procedure in R (R Development Team, 2006) to minimize the RMSD between the predicted and observed response probabilities in each condition.

The resulting functions are plotted as solid lines in Fig. 2A. The average best-fitting (σ, λ) parameter pairs for $Single_{Lower}$ and $Single_{Higher}$ were (1.28, −.04) and (1.91, .01), respectively. Paired comparisons confirmed that while λ did not differ across conditions (and, therefore, did not exceed zero), $t(13) = 1.20$, $p = .25$, σ was clearly greater for $Single_{Higher}$, $t(13) = 6.87$, $p < .001$. As this significant difference in σ does not impact our subsequent analyses or conclusions, which depend on changes in the λ parameter with attention (see below), we reserve discussion of this effect for the General Discussion.

Next, we used the parameters of the best-fitting Skew-Normal distribution to generate predictions for the *Both* condition, assuming simple probability summation of independent visual pattern analyzers (Graham, 1989). Our question is simple: can the mnemometric function obtained for the *Both* condition be described as an additive combination of inputs from independent analyzers tuned to $\mathbf{s}_1$ and $\mathbf{s}_2$? If so, the function generated via probability summation of the observations from the two *Single* distributions should not be exceeded by the observed distribution of *Both* responses between $\mathbf{s}_1$ and $\mathbf{s}_2$. In other words, probability summation formalizes the null hypothesis that perceptual averages do not contribute to the recognition responses observed in the *Both* condition.

[3] These means are prior to normalization.

[4] This effect is not apparent in Fig. 2, which averages across the two $Both_{Lower}$ and $Both_{Higher}$ conditions. Neither the small recency effect in these data, nor the manner in which the data were averaged (aligning based on whether $\mathbf{s}_1$ or $\mathbf{s}_2$ matched the probe, or whether $\mathbf{s}_1$ was Higher or Lower) had any effect on the modeling results reported later in this analysis.

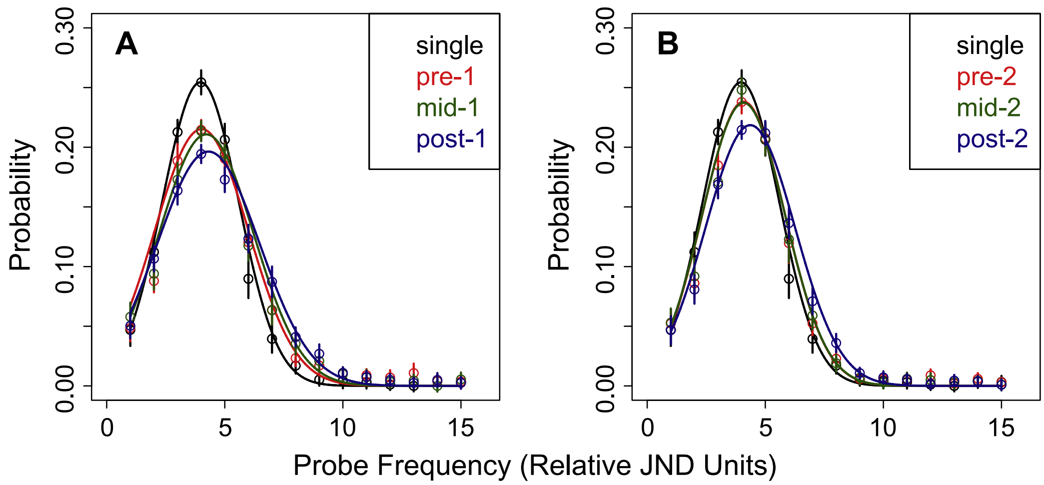

Fig. 3. The mnemometric functions from trials in the *Single*, Pre-Cue, Mid-Cue, and Post-Cue conditions. Shown are the normalized proportions of 'Yes' responses generated as **p** *roves* along the spatial frequency dimension. Observed data are shown as circles, predictions from the best-fitting parameters of the Skew-Normal distribution are shown as solid lines. Error bars are within-subject standard errors (Loftus & Masson, 1994). *Panel A*: data from conditions in which $\mathbf{s}_1$ was designated as the task-relevant study item. *Panel B*: data from conditions in which $\mathbf{s}_2$ was designated as the task-relevant study item.

In the current experiment, stimuli varied only in the vertical spatial frequency dimension. Thus, we assumed the simplest probability summation model, in which summation is based on only one cue (Pirenne, 1943; Treisman, 1998). Furthermore, we chose study items whose spatial frequencies differ by an amount (8 *JNDs*) that is known to minimize interaction among analyzers (Graham, 1989).[5] The distribution of responses to the 15 probes can therefore be described as the sum of the model-predicted vectors corresponding to $Single_{Lower}$ and $Single_{Higher}$, minus the product of predicted probabilities at each probe frequency:

$$\vec{P}_{Both} = \vec{P}_{Single_{Lower}} + \vec{P}_{Single_{Higher}} - \vec{P}_{Single_{Lower}} \circ \vec{P}_{Single_{Higher}} \quad (4)$$

The predictions from probability summation are plotted as Xs in Fig. 2A. The prediction error (Both - Probability Summation) is plotted in Fig. 2B. As is clear in Panel A, the mnemometric distributions for *Both* and the model differ, with inflation of the probability density between the means of $\mathbf{s}_1$ and $\mathbf{s}_2$ evident in the observed *Both* data. Panel B shows that 95% CIs of the difference at each point support this interpretation.

Note that there are, however, two large deviations at 4 and 5 *JNDs* in both panels. This is likely due to the large difference in σ for the two *Single* distributions. Though the origin of this difference in the *Single* distributions is unclear, the change in σ cannot account for the systematic differences observed in the region of the average, which are nonetheless replicated in analyses that do not depend on the σ parameter, to which we now turn.

3.2. Selective attention conditions

We now consider the selective attention conditions, in order to address our second hypothesis, i.e. that subjects will show a reduced influence of perceptual averaging when selective attention is directed to the relevant study item in advance. The two panels of Fig. 3 show the mnemometric functions produced when selective attention is directed to $\mathbf{s}_1$ (Panel A) or $\mathbf{s}_2$ (Panel B), and for the different cue timepoints in a trial. The data from the baseline condition, *Single*, are replotted for comparison to the selective attention conditions. All distributions are plotted against relative *JND* units: 0 is the condition of minimum *JNDs* (i.e. 0) for the distribution whose target value is 4 *JNDs*, but is the condition of maximum *JNDs* (i.e. 15) for the distribution whose target value is 12 *JNDs*. Expressing the distributions in this manner allows for a straightforward interpretation of changes in their form: increasing (positive) skew indicates inflation of probability mass toward the perceptual average of each trial's study items.

Differences in shape across the mnemometric functions within each panel suggest that cue timing was not without effect. Specifically, as the presentation of the cue is increasingly delayed, the mnemometric function is increasingly skewed toward the irrelevant study item. This inflation corresponds to an increase in false alarm rates in the region spanning the perceptual average of $\mathbf{s}_1$ and $\mathbf{s}_2$ on each trial (i.e. 5–10 *JNDs*, each panel). To capture what might be an important global effect, examining P (*Yes*) at one or at just a few **p** values would be inadequate. Instead, to generate an appropriate, scalar distributional measure for each mnemometric function we again fit the Skew-Normal, as in our previous analysis of the *Single* condition.

The resulting skewness values (λ_{corr}) were sign-corrected, such that skew toward the average is positive and away is negative. The parameter values for individual subjects, along with the corresponding σ values, are shown in Table 1; group-averaged values of λ_{corr} are shown in Fig. 4. A repeated measures ANOVA confirmed that the increase in skewness relative to *Single* (baseline-subtracted λ) varied significantly as a function of cue timing, $F(2,26) = 3.96$, $p < .05$. Neither the main effect of recency (whether $\mathbf{s}_1$ or $\mathbf{s}_2$ was the relevant study item) nor its interaction with cue timing were significant (all $ps > .05$). Paired comparisons showed that the increase in skewness was greater in the Post- than in the Pre-Cue condition, $t(11) = 2.46$, $p < .05$, and was also greater in the Post- than in the Mid-Cue condition, $t(11) = 2.86$, $p < .05$. This result suggests an increased influence of perceptual averages as effective deployment of selective attention is delayed.

The σ parameter showed greater spread in the $\mathbf{s}_1$ than in the $\mathbf{s}_2$ condition, $F(1,13) = 27.68$, $p < .001$, and a marginally significant effect of cue timing, $F(2,26) = 3.34$, $p = .051$. In the latter case, paired comparisons showed that the only significant effect was that σ was greater in the Post- than in the Mid-Cue condition, $t(11) = 2.48$, $p < .05$. These effects are generally consistent with weaker and/or noisier memory representations in conditions of greater lag (i.e. time elapsed between two stimuli, being greatest for $\mathbf{s}_1$ in this case) or memory load (Post-Cue).

[5] Our modeling assumptions about independence, which apply only to the probability summation analysis of the Both condition, are based on results from threshold conditions with simultaneously-presented gratings, conditions that differ somewhat from our own (Graham, 2011).

Table 1
Best-fitting parameters of the skew-normal distribution. Reported are parameters from fits to individual subjects, the average of those parameter values, and parameters from the fits to group average data.

ID	λ_{Corr}							σ						
	Single	Pre-1	Mid-1	Post-1	Pre-2	Mid-2	Post-2	Single	Pre-1	Mid-1	Post-1	Pre-2	Mid-2	Post-2
1	0.24	0.09	−0.11	0.19	0.13	0.31	0.24	1.54	2.04	1.94	2.30	1.77	1.65	1.73
2	−0.12	−0.04	−0.15	0.16	0.16	0.16	0.18	1.20	1.55	1.05	2.06	1.37	1.30	1.44
3	−0.13	−0.42	0.08	−0.06	0.16	−0.02	0.50	1.48	1.95	1.58	2.22	1.34	1.38	1.92
4	−0.18	−0.08	0.03	0.06	−0.03	−0.31	0.30	1.73	1.78	1.91	2.11	1.52	1.74	1.99
5	0.03	0.13	−0.12	−0.01	0.09	0.23	0.49	1.70	1.43	1.77	2.36	1.71	1.51	1.93
6	0.06	0.34	0.46	0.56	0.06	0.12	0.52	1.82	1.87	2.22	2.05	2.05	1.80	2.05
7	−0.66	0.19	−0.38	−0.03	−0.22	−0.27	−0.01	1.57	2.56	3.21	2.48	1.97	1.87	2.45
8	−0.26	0.32	0.05	0.38	0.14	0.13	0.12	1.88	2.11	2.13	2.38	1.89	1.79	2.45
9	−0.25	0.11	0.06	−0.30	0.13	−0.15	0.05	2.18	1.84	1.74	2.17	1.50	1.58	1.62
10	0.33	1.21	0.78	1.57	0.56	0.29	0.44	1.45	2.01	1.67	2.42	2.01	1.69	1.85
11	−0.25	0.06	0.46	0.22	0.12	0.75	0.47	2.06	2.94	2.75	1.42	2.29	1.97	2.13
12	0.22	−0.13	0.39	0.35	0.07	0.20	0.09	2.02	4.24	3.05	3.81	2.98	2.56	2.92
13	−0.10	0.16	0.23	−0.04	0.24	−0.34	0.05	1.38	1.65	1.88	1.71	1.65	1.32	1.34
14	−0.07	0.08	0.06	0.15	0.24	0.22	0.56	1.11	1.38	1.33	1.38	1.17	1.23	1.16
Average	−0.08	0.14	0.13	0.23	0.13	0.09	0.28	1.65	2.10	2.02	2.21	1.80	1.67	1.93
Group	−0.02	0.02	0.14	0.18	0.06	0.06	0.25	1.62	1.98	2.01	2.19	1.74	1.75	1.93

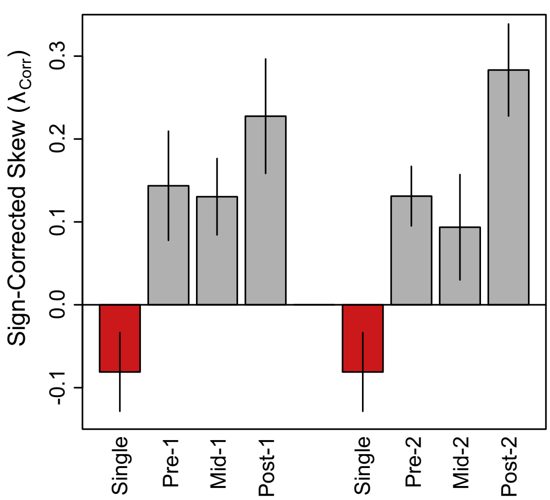

Fig. 4. Mean skew (sign-corrected λ) associated with condition *Single* and the six conditions of selective attention. Each parameter value is an average over individual-subject parameter values. The bar representing *Single* is shown twice in the figure in order to facilitate comparisons involving conditions in which $\mathbf{s}_1$ was task-relevant (left side) and conditions in which $\mathbf{s}_2$ was task-relevant (right side).

4. Discussion

Huang and Sekuler (2010a) demonstrated that irrelevant items influence VSTM responses (see also Huang & Sekuler, 2010b). Specifically, reproduction of remembered spatial frequency is shifted in the direction of the irrelevant item's spatial frequency. To account for this result, Huang and Sekuler (2010a) proposed a model that assumes the effect operates at an early, encoding-related stage of VSTM. They hypothesized that subjects stored a weighted average of a trial's relevant and irrelevant memory items, which influenced subsequent recall in the absence of selective attention. We tested this hypothesis further, by fixing the similarity of individual study items and varying the spatial frequency match of a recognition probe. This roving probe technique allowed us to construct mnemometric functions, to which we applied a distributional analysis. This fine-grained technique allowed us to more clearly characterize the nature of the effect reported originally by Huang and Sekuler (2010a).

On each trial of our experiment, we delivered cues as to which of two study items was to be matched to the upcoming test probe. Importantly, this cue appeared either before (Pre-Cue condition), between (Mid-Cue condition), or after (Post-Cue condition) the two study items of a given trial. Analyses of mnemometric functions showed that, in the Post-Cue conditions, lures spanning the spatial frequency range of the average of the study items produced elevated false alarm rates. This pattern was less pronounced in the Pre-Cue conditions, suggesting selective attention can be used to improve encoding of target items. Such improved encoding should reduce subjects' uncertainty regarding VSTM's contents, as well as any need to compensate for such uncertainty (Ball & Sekuler, 1980).

Though the effects of our attentional manipulation are clear, there were changes in variance of the *Single* distributions, the origin of which is not clear. Specifically, σ was higher for $Single_{Higher}$ than for $Single_{Lower}$. Study-test lag was equated in this condition, which involved only one study item always presented at the same timepoint in a trial. As the only factor that varies is spatial frequency of the **s**, it seems possible that higher spatial frequencies (which are higher above the detection threshold and, in a sense, 'stronger') produce more variable distributions of memory strength. From the standpoint of long-term recognition memory, this makes sense, as successful memory strength models in that domain typically assume that representations with higher strength are also more variable (Dubé et al., 2013; Wixted, 2007). Though future work will be necessary to understand the nature of this difference, we see no obvious way in which the frequency effect in σ could have produced the systematic effects of similarity, attention, and memory load that we observed in this experiment, especially as the critical effects we reported all involved changes in λ.

Our results also failed to show a strong effect in the Mid-Cue condition, which produced a distribution of responses similar to that of the Pre-Cue condition, and no significant skew. However, it is not obvious that the Pre- and Mid-Cue conditions should be expected to show large differences to begin with. That is, the Mid-Cue condition allows one of two possible advantages on a given trial: either subjects can ignore an upcoming irrelevant item ($\mathbf{s}_1$ trials) or they can prepare to focus attention on an upcoming relevant item ($\mathbf{s}_2$ trials), but they cannot do both. There is thus a partial benefit on any Mid-Cue trial, while in the Post-Cue condition, neither opportunity presents itself; irrelevant items are equally likely to be attended, and subjects are unable to devote full attention to the task-relevant study item. If only one of the two scenarios provided by the Pre-Cue condition is enough to override uncertainty in VSTM's fidelity, then one would not expect a difference between the Pre- and Mid-Cue conditions. Unfortunately, the present data do not allow us to do much more than speculate on the reasons for the convergence between the Pre- and Mid-Cue conditions. However, we also do not see this convergence as posing a strong challenge to our results or conclusions.

As a final caveat, we note that our results do not allow strong inferences about the mechanism by which averages influence recognition responses. It is clear that subjects, even in conditions of total ambiguity as to the relevant stimulus, do not sacrifice their representations of $\mathbf{s}_1$ and $\mathbf{s}_2$ when making recognition responses, and show low false alarm rates to **L**s regardless of their spatial frequencies. This suggest that, if subjects are computing and storing perceptual averages, they are maintaining at least three memory representations on a given trial. Furthermore, if responses reflect a weighted sum of spatial frequency matches (as is assumed in global matching models such as NEMo; Sekuler & Kahana, 2007), the weight assigned to matches against the perceptual average is likely exceeded by the weights for matches to representations of $\mathbf{s}_1$ and $\mathbf{s}_2$.

On the other hand, the present study does carry clear implications for models of VSTM that assume a fixed number of independent "slots" for memory representations (Lee & Chun, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008, 2009, 2011). Our findings suggest that successive incoming stimuli interact during encoding, with each stimulus' representation contributing to the computation of a perceptual average, and are not insulated via entry into separate memory slots. Although further speculation clearly requires extrapolation beyond our data, our results can easily be construed as supportive of flexible-resource models (Bays & Husain, 2008; Bays, Catalao, & Husain, 2011; Wilken & Ma, 2004), in which potential targets must compete for a shared resource at encoding. If this account is true, then reliance on perceptual averages may be a way for VSTM to resolve such competition. This suggests future work should examine how manipulations of selective attention affect performance measures in tasks that have been used to differentiate between VSTM models (see, e.g., Zhang & Luck, 2009).

Our findings show that subjects compute perceptual averages within each trial. Such effects are distinct from ones that result from averaging across trials (Huang & Sekuler, 2010a).[6] However, the existence of these two effects seems to require that subjects compute and maintain separate representations of the two averages. How are subjects able to differentiate the intervals over which the two averages are computed?

A recent study by Lohnas, Polyn, and Kahana (in preparation) suggests an answer. Those authors proposed a model of free recall that they referred to as the modified Context Maintenance and Retrieval Model (CMR2). They showed that the CMR2 provides a principled account of how subjects in a recall experiment can separately direct memory search to items within and across lists. CMR2 accomplishes this by assuming that events which signal a shift in test context (e.g., a recall trial between two study trials) produce distinct temporal contexts (e.g., the two study trials) to which the items that were presented at those times are associated. In our study, this would imply that each recognition event may help to insulate the current trial from previous trials, allowing computation of a perceptual average that produces effects that are distinct from across-trial effects (Huang & Sekuler, 2010a).

Though the effects of perceptual averaging in our study are clear, it is also clear that averages are not the only statistics that subjects consider in VSTM tasks such as ours. For instance, several studies have demonstrated that the similarity of individual study items to one another (i.e. their homogeneity) influences subjects' willingness to make a 'Yes' response to a recognition probe (i.e. response bias). All else being equal, subjects make fewer 'Yes' responses when study item homogeneity is greater (Viswanathan et al., 2010). This suggests that the variance in items' features is also an important determinant of recognition responses. Furthermore, it has been shown that subjects' memory for serial position of individual study items is more accurate when when homogeneity of the study items is lower (Yotsumoto et al., 2008).

Although most studies of visual recognition memory focus on discrete measures such as hits and false alarms, our focus encompassed more global assays of recognition, in particular, the distribution of false alarms across several levels of inter-item similarity. We believe that these mnemometric functions (Sekuler & Kahana, 2007) provide a powerful window onto recognition memory. In the present case, they are particularly informative as to the possible origin of a false recognition in conditions involving selective attention. For example, confusion of the cued- or non-cued status of the two study items could potentially have led subjects to respond "Yes" when the **p** matched the to-be-ignored stimulus. In other words, a portion of what seem to be false recognitions could have arisen had the subject failed to comply with the instruction to attend to just one of the study items. This possibility is ruled out by the mnemometric functions of the selective attention conditions. Specifically, for all conditions of selective attention, the mnemometric function approached or reached its lower limit as **p** deviated from the attended stimulus, demonstrating that recognition decisions were strongly influenced by the similarity of **p** to the relevant study item. More importantly, the false alarm rates in selective attention conditions were consistently at the lower limit when **p** matched the task-irrelevant study item, as can be seen in Fig. 3's mnemometric functions. One-sample t-tests failed to reveal any response rates above zero for probes matching the irrelevant study item, largest $t(11) = 1.77,\ p = .10$ (*Pre-2* condition). The fact that subjects made essentially zero false alarms when **p** matched the task-irrelevant study item, but showed inflated false alarm rates near the average of the two stimuli, suggests that our results reflect the influence of perceptual averages, rather than an influence of the irrelevant item's representation per se.

5. Conclusions

Our results suggest three important facts about perceptual averaging. First, perceptual averages (at least in our task) are computed in the absence of any explicit requirement to do so. This builds on previous work examining perceptual averaging, as such studies have typically required subjects to report an estimate of the perceptual average of a given trial (Alvarez, 2011). Second, perceptual averages appear to be used as compensatory representations. That is, under conditions where memory representations are weaker, subjects' VSTM responses are more susceptible to the average's influence. This makes sense given previous work which has shown (i) that subjects rely on perceptual averages under conditions of stimulus uncertainty (Ball & Sekuler, 1980; Zanto et al., 2013) and (ii) that perceptual averages are relatively robust to division of attention at encoding (Alvarez & Oliva, 2008, 2009). Third, our study supports previous claims that perceptual averages are computed over time as well as space (Albrecht & Scholl, 2010; Haberman, Harp, & Whitney, 2009).

Acknowledgments

Research supported by National Institutes of Health Grants MH55687, MH68404 and AFOSR Grant FA9550-10-1-0420. Some of this report is based on data collected as part of F.Z.'s

[6] An analysis of Yes rates as a function of Target probe similarity to the average stimulus over all trials in the Post-Cue condition failed to detect any effect of the long-range average in the present experiment. This further supports our interpretation of the within-trial effect as a perceptual averaging phenomenon.

Ph.D. dissertation work at Brandeis University; he is now with Weaver, Austin, Velleneuve, and Sampson, LLP, Oakland.

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.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention Perception and Psychophysics, 74*, 810–815.

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 of the United States of America, 106*, 7345–7350.

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

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandanavian Journal of Statistics, 12*, 171–178.

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.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20*, 160–166.

Bansal, N. K., Maadooliat, M., & Wang, X. (2008). Empirical Bayes and hierarchical Bayes estimation of skew normal populations. *Communications in Statistics: Theory and Methods, 37*, 1024–1037.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2011). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*, 698–711.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851–854.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

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

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.

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

Graham, N. (2011). Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision Research, 51*, 1397–1430.

Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expressions over time. *Journal of Vision, 9*, 1–13.

Huang, J., & Sekuler, R. (2010a). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*, 24.1–27.

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

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

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model. *Memory & Cognition, 35*, 1222–1232.

King-Smith, P., Grigsby, S., Vingrys, A., Benes, S., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*, 885–912.

Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception and Psychophysics, 63*, 253–257.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2014). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall (submited for publication).

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences, 2*, 187–199.

Pirenne, M. H. (1943). Binocular and uniocular threshold of vision. *Nature, 152*, 698–699.

R Development Team (2006). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin.

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

Treisman, M. (1998). Combinding information: Probability summation and probability averaging in detection and discrimination. *Psychological Methods, 3*, 252–265.

Viswanathan, S., Perl, D. R., Visscher, K. M., Kahana, M. J., & Sekuler, R. (2010). Homogeneity computation: How inter-item similarity in visual short term memory alters recognition. *Psychonomic Bulletin & Review, 17*, 59–65.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*, 113–120.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120–1135.

Williams, H. D., Titchener, E. B., & Boring, E. G. (1918). On the calculation of an associative limen. *American Journal of Psycholology, 29*, 219–226.

Wixted, J. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152–176.

Yotsumoto, Y., Kahana, M. J., McLaughlin, C., & Sekuler, R. (2008). Recognition and position information in working memory for visual textures. *Memory & Cognition, 36*, 282–294.

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.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.

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

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*, 1434–1441.

Zhou, F., Kahana, M. J., & Sekuler, R. (2004). Short-term episodic memory for visual textures: A roving probe gathers some memory. *Psychological Science, 15*, 112–118.