

Homogeneity computation: How inter-item similarity in visual short term memory alters recognition

Shivakumar Viswanathan[†], Daniel R. Perl,
Kristina M. Visscher^{*}, Michael J. Kahana[‡] and Robert Sekuler

Volen Center for Complex Systems, Brandeis University

[†] Department of Psychology, University of California, Santa Barbara

^{*} Department of Neurobiology, University of Alabama, Birmingham

[‡] Department of Psychology, University of Pennsylvania

Visual short-term recognition memory for a sequence of stimuli is strongly influenced by the study item's similarity to one another, that is, by their homogeneity. However, the mechanism responsible for this homogeneity effect has remained unclear. We evaluated competing explanations of this effect, using controlled sets of Gabor patches as study items and probe stimuli. Our results, based on recognition memory for spatial frequency, rule out the possibility that the homogeneity effect arises because similar study items are encoded and/or maintained with higher fidelity in memory than dissimilar study items. Instead, our results support the hypothesis that the homogeneity effect reflects trial-by-trial comparisons of study items to one another, which generate a homogeneity signal. This homogeneity signal modulates recognition performance through an adjustment of the subject's decision criterion. Additionally, it seems the homogeneity signal is computed prior to the presentation of the probe stimulus, by evaluating the familiarity of each new stimulus with respect to the items already in memory. This suggests that recognition-like processes operate on not only the probe stimulus, but on study items as well.

Visual short term memory (VSTM) actively maintains information about stimuli that have recently disappeared from view. A new incoming stimulus can automatically interact with items already in VSTM. These interactions are sensitive to the new stimulus' similarity to the items in memory, and can occur even when the new stimulus is task irrelevant (Miller & Desimone, 1994; Grill-Spector, Henson, & Martin, 2006; Magnussen, 2000; ?). Here we tested competing accounts of how the similarity between multiple, sequentially-presented items influences recognition performance.

Our test used a variant of Sternberg's recognition paradigm (1966). On each trial, a subject saw a sequence **S** of multiple stimuli (study items). Then, following a brief delay, a single probe stimulus **p** was presented, and the subject judged whether **p** replicated one of the studied items in **S**, responding YES if this was the case or NO otherwise. As stimuli in **S** as well as **p** varied across trials, subjects had to maintain each of that trial's items in memory, and then compare **p** to these remembered items in order to make a recognition judgment.

The similarity of the probe to each of the study items

strongly influences recognition, a phenomenon that is well explained by Global Matching Models (Nosofsky, 1991; Clark & Gronlund, 1996; Zaki & Nosofsky, 2001; Lamberts, Brockdorff, & Heit, 2003). Such models postulate a global matching process whereby the probe, **p**, is compared to the memory representation of each study item, with each comparison yielding a scalar similarity signal. These separate signals are combined into a single familiarity signal, which is compared against a decision criterion in order to produce a recognition judgment. These models predict that the probability of a YES response (hereafter, $P(\text{YES})$) will tend to be higher when **p** is simultaneously similar to multiple study items rather than to just one study item.

Studies of VSTM with the Sternberg paradigm have revealed another, independent effect of stimulus similarity that putatively reflects the similarity of the study items to one another (Kahana & Sekuler, 2002). Specifically, when the study items in VSTM are similar to one another (i.e., are "homogeneous"), subjects tend to make *fewer* false recognitions than predicted by standard similarity-based recognition models. This effect of study-item similarity, which we will refer to as the *homogeneity effect*, has been confirmed with various stimuli, both visual (Kahana & Sekuler, 2002; Nosofsky & Kantner, 2006; Kahana, Zhou, Geller, & Sekuler, 2007; Yotsumoto, Kahana, Wilson, & Sekuler, 2007) and auditory (Visscher, Kaplan, Kahana, & Sekuler, 2007), and has been subjected to detailed, model-based analysis (Kahana & Sekuler, 2002; Nosofsky & Kantner, 2006; Visscher et al., 2007; Kahana et al., 2007). Despite the atten-

Research supported by NIH grant MH068404 and MH55687; e-mail: shiva@brandeis.edu or sekuler@brandeis.edu. We thank Robert Nosofsky, the anonymous reviewers and Heather Sternschein for valuable comments.

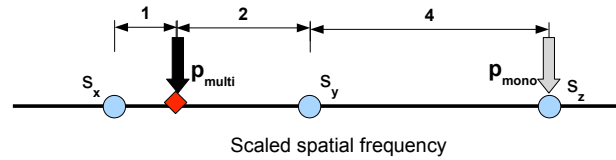
tion the homogeneity effect has attracted, the mechanism behind the effect has remained unclear. In this study, we empirically evaluated two competing explanations of the homogeneity effect.

The first of these possible explanations was suggested by behavioral and physiological evidence that the similarity of sequentially-presented stimuli systematically influences the fidelity with which the stimuli are represented in memory (Spitzer, Desimone, & Moran, 1988; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Bennett & Cortese, 1996; Magnussen, 2000). According to this hypothesis, study items whose feature values are similar to one another are each maintained in memory with higher fidelity than dissimilar study items would be. The heightened fidelity of memory representations would reduce the likelihood of a false recognition when the study items were highly similar to each other. We refer to this account of the homogeneity effect as the *Memory-Precision* hypothesis.

A second hypothesis, proposed by Kahana and Sekuler (2002), asserts that the familiarity signal postulated by Global Matching Models is supplemented by a second, homogeneity-dependent signal; and that recognition depends on both signals. To generate this second signal, scalar similarity values are obtained from pairwise comparisons of the study items to one another. These inter-item similarity values are then averaged to produce a scalar measure of homogeneity, which represents the degree of similarity of the study items in **S** to one another. This homogeneity signal subsequently influences the recognition judgment by modulating the familiarity signal. Nosofsky and Kantner (2006) proposed an alternative, namely, that the homogeneity signal is used to adaptively to adjust the decision criterion. This adjustment would offset the reduced accuracy in rejecting lures that Global Matching Models predict for highly homogeneous study lists. As both these interpretations impute an independent computation of study set homogeneity, we refer to these two possibilities together as the *Homogeneity Computation* hypothesis.

To select between these competing hypotheses, we used the Sternberg paradigm with lists of three study items. We identified critical configurations of study items and probes for which the two competing hypotheses make conflicting predictions. This allowed us to directly select between these hypotheses. The configurations we identified and the predictions of the two hypotheses are described next. As our empirical test required careful control over the similarity of each trial's study items and their associated probes, we used Gabors – vertical sinusoidal luminance gratings windowed by a circular Gaussian – that could vary in spatial frequency. The metric properties of these stimuli provided a common measure to compare trials, and stimuli could be adjusted to control for differences in subjects' perceptual performance (Zhou, Kahana, & Sekuler, 2004).

Study list **S**



Study list **S'**

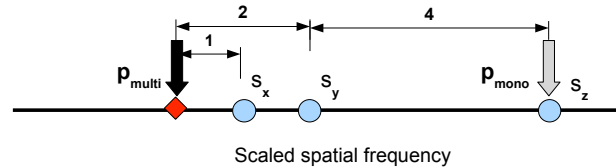


Figure 1. The study lists **S** (upper panel) and **S'** (lower panel) differ in their degree of homogeneity. For both lists, the probe \mathbf{p}_{multi} is similar to s_x and to s_y , but \mathbf{p}_{mono} is highly similar to s_z while being far less similar to the other study items. Note that the similarity of \mathbf{p}_{multi} and \mathbf{p}_{mono} to each of the study items is maintained on both lists, for reasons explained in the text.

Experiment design and Predictions

Figure 1 depicts schematically the design of the study lists and probe stimuli. The figure's panels illustrate the relative spatial frequencies of the study items and the critical probes of interest for two types of study lists, **S** (upper panel) and **S'** (lower panel). In each panel, the thick horizontal line represents spatial frequency scaled to units that are integer multiples of each subject's discrimination threshold or JND (see the Methods section below for details of the scaling procedure). The three discs on each horizontal line represent the spatial frequencies of a trial's study items, which we label s_x , s_y and s_z . The thick vertical arrows indicate the spatial frequencies of the critical probes in relation to the study items. As is customary, we use the term "target" for a probe that matches one of the study items, and the term "lure" for a probe that does not. In Figure 1, a lure is represented by a diamond.

For purposes of describing the experiment design, we treat the absolute difference in scaled spatial frequency between two stimuli on this one-dimensional continuum as a measure of the perceptual similarity of these two stimuli via a non-linear monotonic relationship (for example, see Shepard, 1987). Note that the degree of similarity among study items in list **S'** is greater than in **S** because the absolute difference (or distance) between s_x and s_y is smaller in **S'** than in **S**. Consequently, the baseline prediction is that the homogeneity effect should have a larger influence on recognition judgments with study lists of type **S'** than on lists of type **S**. To measure how recognition judgments are influenced by this difference in the two lists' homogeneity values, we used two types of probes, referred to as **MULTI**-probes and **MONO**-probes. These are described next.

\mathbf{p}_{multi} , the probe represented by the black arrows in

both panels is a lure that is one JND away from s_x , two JNDs from s_y and six JNDs from s_z . This probe is referred to as a MULTI-probe as, in both types of study lists, it has a relatively small distance to both s_x and s_y , i.e., it is simultaneously similar to multiple study items. \mathbf{p}_{mono} , the probe indicated by the gray arrow in both panels, is a target probe that replicates study stimulus s_z . This probe is referred to as a MONO-probe as it is highly similar to one of the study items, but is dissimilar to the other study items. As Figure 1 shows, the distance of \mathbf{p}_{mono} to item s_z is zero, but its distance to the other two study items is relatively “large”, i.e. at least four JNDs.

These inter-stimulus distances were chosen to ensure that \mathbf{p}_{multi} 's distance to each study item in \mathbf{S} would be preserved in the other type of list, \mathbf{S}' (also see Visscher et al., 2007). This constraint is approximately true for \mathbf{p}_{mono} as well. In both \mathbf{S} and \mathbf{S}' list types, \mathbf{p}_{mono} is equally similar to s_y and s_z , but is extremely dissimilar from s_x (seven and five JND units different, respectively). So, in terms of perceptual similarity, \mathbf{p}_{mono} is equally similar to s_x on lists \mathbf{S} and \mathbf{S}' . Thus the two list types are very different in their study items' homogeneity, but the respective distances between the probes and the study items are equivalent on both study lists. As a result, any differences in recognition performance between the two study lists would be attributable to the difference in the study lists' homogeneity. The competing hypotheses introduced earlier make conflicting predictions on exactly how the recognition judgments might differ on the probes \mathbf{p}_{multi} and \mathbf{p}_{mono} . These predictions, which are the focus of our data analysis, are presented below. The probability that subjects respond YES on a particular condition ($P(\text{YES})$) is the dependent variable.

The *Memory-Precision* hypothesis asserts that the noise associated with study items' memory representations is reduced when study items are highly similar to each other. If this were the case, the variance in the memory representations of s_x and s_y on list \mathbf{S}' should be lower than for the corresponding study items on list \mathbf{S} .

The behavioral consequence of this “sharpening” in the memory representations with increased homogeneity is that lures should be endorsed less often on \mathbf{S}' as compared to \mathbf{S} . Therefore, for the lure probe \mathbf{p}_{multi} , this hypothesis predicts that

$$P(\text{YES} | \mathbf{p}_{multi}, \mathbf{S}') < P(\text{YES} | \mathbf{p}_{multi}, \mathbf{S}). \quad (1)$$

Additionally, because memory representations for s_x and s_y are “sharper,” targets should be endorsed more often on \mathbf{S}' as compared to \mathbf{S} . This hypothesis predicts no difference in the memory representation of s_z between the two lists, because study-item s_z is dissimilar to s_x and s_y on both \mathbf{S} and \mathbf{S}' . However, there may be some relatively small reduction in noise as s_z is somewhat closer to s_x on \mathbf{S}' as compared to \mathbf{S} . Therefore, for the target probe, \mathbf{p}_{mono} , the prediction is that

$$P(\text{YES} | \mathbf{p}_{mono}, \mathbf{S}') \geq P(\text{YES} | \mathbf{p}_{mono}, \mathbf{S}) \quad (2)$$

Thus, the *Memory-Precision hypothesis* predicts that if $P(\text{YES})$ did differ between \mathbf{S} and \mathbf{S}' , their differences should be of opposite sign for the two probe types.

The *Homogeneity-Computation hypothesis* asserts that a computation of study-list homogeneity influences recognition judgments on *all* probes, and does so independently of the degree of similarity of the probe to the study items. Therefore, this hypothesis predicts that the change in $P(\text{YES})$ value on both \mathbf{p}_{multi} and \mathbf{p}_{mono} between the two lists would have the *same* sign, that is,

$$P(\text{YES} | \mathbf{p}_{multi}, \mathbf{S}') < P(\text{YES} | \mathbf{p}_{multi}, \mathbf{S}) \quad (3a)$$

$$P(\text{YES} | \mathbf{p}_{mono}, \mathbf{S}') < P(\text{YES} | \mathbf{p}_{mono}, \mathbf{S}) \quad (3b)$$

Methods

Stimuli. Each Gabor stimulus subtended 5.6° at a viewing distance of 114 cm. A Gabor's mean luminance was 50 cd/m^2 , and its sinusoidal component had a peak contrast of 0.20. Different stimuli were generated by varying f , the spatial frequency of the Gabor's sinusoidal component (described below). On each presentation of a stimulus, the phase of its sinusoidal component was varied randomly over the range $[0, \pi/2]$, which forced subjects to make judgments on spatial frequency rather than on any local, retinotopic detail. Stimuli were generated and displayed using Matlab and Psychtoolbox (Brainard, 1997) on a 32 cm x 24 cm CRT monitor with a screen resolution of 1152×864 pixels.

Stimulus scaling. The stimulus set for each subject was generated by a subject-specific scaling procedure (Zhou et al., 2004). A subject's stimulus set consisted of spatial frequencies defined by the relation $f = f_0(1 + K_{subject})^n$ where f_0 is a fixed base frequency. $K_{subject}$ was the subject's own Weber fraction providing an estimate of the smallest difference in spatial frequency that the subject discriminates correctly 85% of the time (i.e., the Just Noticeable Difference or JND). The variable n is an integer that defines the difference between f and f_0 in JND units. In our experiment, n assumed integer values in the range $[-6, +7]$. This defined a set of 14 stimuli for which spatial frequencies in stimulus pairs differed by an integer number of JND units. The base frequency was set to $f_0 = 1.43 \text{ cycles/degree}$. To prevent subjects from memorizing these individual stimuli, a second set of 14 stimuli was generated with a slightly different base frequency obtained by incrementing f_0 by 0.5 JNDs. This “jittered” stimulus set was used on half the trials, chosen randomly. All stimuli on a particular trial were drawn from only one of these two stimulus sets. Our data analysis aggregated trials from the two stimulus sets.

Study lists. Three types of study lists, each with three items, were used. Each list type was defined only by the absolute distances (in JND units) between study items. The two list types needed to select between the competing hypotheses are shown graphically in Figure 1. A third list type was introduced to keep subjects from adopting a strategy

specifically tuned to these two lists. In this third list, the distance between s_x and s_y was constrained to be 4 JNDs, while the distance between s_y and s_z was 4 JNDs, as in the two lists shown in Figure 1.

As the changes in the distance between s_x and s_y governs the difference in the homogeneity of these three lists, we adopt the following nomenclature. Lists where the distance between s_x and s_y was 1 JND, i.e. the highly homogenous lists, will be referred to as the HIGHHOM type; when 3JNDs as the MEDHOM type; and when 4 JNDs as the LOWHOM type. The two study lists **S** and **S'** shown in Figure 1 correspond to the MEDHOM and HIGHHOM list types respectively. We will also refer to s_y as the MIDDLE study item as it lies between the other two study items; s_x as the CLOSE study item due to its variable distance to MIDDLE; and s_z as the FAR study item due to its larger (and constant) distance of 4 JNDs to MIDDLE.

Individual lists for each of the list types was defined only by the absolute distances (in JND units) between the study items, as described above. We generated every triple of spatial frequencies that satisfied these list-specific constraints. These spatial frequency triples occurred with equal probability on trials of each of the list types. This procedure ensured that there was no preferred spatial frequency relationship between the CLOSE, MIDDLE and FAR study items. For example, one list of the HIGHHOM type could be such that $f_{Close} < f_{Middle} < f_{Far}$, where f_i is the spatial frequency of stimulus i , and another could be such that $f_{Close} > f_{Middle} > f_{Far}$. Furthermore, a stimulus having a particular spatial frequency could not be used to predict the list-type being tested. The sequential presentation order of the CLOSE, MIDDLE and FAR study items for each list type was randomized, with each of the six possible unique presentation orders being equally likely.

Probes. Each subject performed 1620 trials (50% target trials, 50% lure trials). An equal number of trials (540) were devoted to each of the three list types, HIGHHOM, MEDHOM and LOWHOM. Within each type of list, target and lure trials occurred with equal frequency. For each list type, the target probes matched the study item at each of the three serial position on one third of the target trials. The target matching the FAR study item on lists HIGHHOM and MEDHOM was the critical probe \mathbf{p}_{mono} . The lure set contained the critical probe \mathbf{p}_{multi} . To prevent subjects from overtly using the perceived similarity of the study items as a cue to predict the “difficulty” in judging the probe, we ensured that the lure trials for each list type were (approximately) equivalent in difficulty. The set of possible lures were constrained to always lie within the range $s_{low} - 3$ and $s_{high} + 3$ on each list, where s_{low} is the study item having the lowest spatial frequency on a trial, and s_{high} is the one with higher spatial frequency. The lure-set for each list type was divided equally into two groups – “hard” and “easy”. Lures in the “hard” group were 1 JND away from the nearest study-item on the list, and lures in the “easy” group had a distance > 1 JND. Lure trials were split equally between these two groups. On list types MEDHOM and HIGHHOM, the “hard”

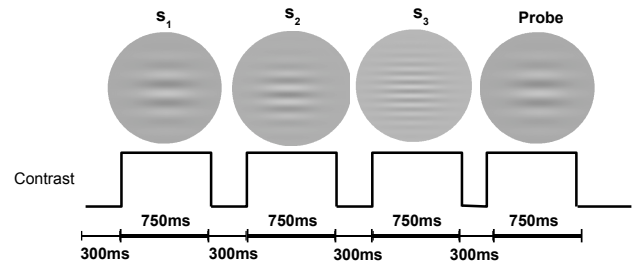


Figure 2. The sequence and timing of events on a trial. On each trial of the experimental task, subjects first fixated on a “+” at the center of a computer screen for 750 ms. After this, each of the study-items was presented for 750 ms with a 300ms interval during which a blank screen was presented. After the last item from a list was presented, a short beep sounded, and an ellipsis “...” was displayed for a 300 ms retention interval, indicating that the subject should wait for the probe. This was followed by the display of the probe for 750 ms. The trial ended when the subject responded by pressing either of two pre-designated keys. Feedback was provided using distinctive tones that indicated whether their response had been correct or incorrect. Between trials, the screen indicated the percent correct trials.

lures contained the MULTI-probe \mathbf{p}_{multi} . This lure was presented on a total of 90 trials for each of these two list-types.

The ordering of all list types across trials was randomized, and for each type, an approximately equal number of trials were presented in each of the five sessions required to complete the experiment.

Subjects. Ten subjects (three male, seven female) recruited from the Brandeis student population participated in the experiment. All subjects were paid and were between 18 and 23 years old (mean = 20 y). The experiment comprised five sessions of about 50 minutes each. Successive sessions were separated by at least three hours; all sessions were completed within two weeks.

Procedure. Before the first experimental session, subjects underwent a vision screening that ensured their Snellen acuity was normal or corrected-to-normal. After this screening, each subject’s Weber fraction for spatial frequency was estimated using an adaptive psychophysical procedure (Watson & Pelli, 1983). Figure 2 summarizes the sequence and timing of events on a trial in the main experiment. Trials were self-paced, with subjects pressing a key to start a trial. Subjects received 30 practice trials prior to each session, and were instructed to be accurate and quick with their responses.

Results

Figure 3 shows the mean P(YES) values with the MULTI-probe \mathbf{p}_{multi} and the MONO-probe \mathbf{p}_{mono} , on the MEDHOM and HIGHHOM lists. For the probe \mathbf{p}_{multi} , subjects were significantly less likely to respond YES on the HIGHHOM list

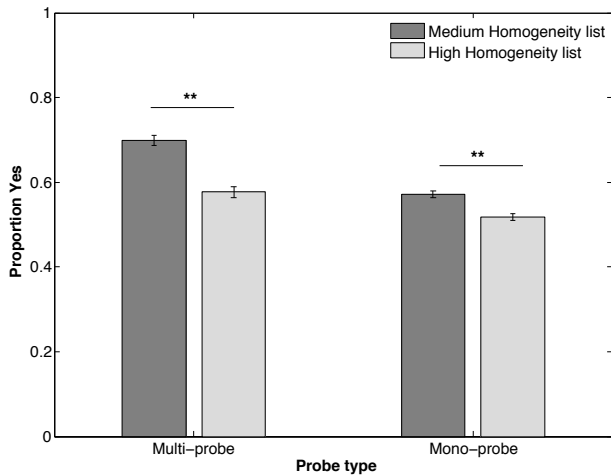


Figure 3. P(YES) values for \mathbf{p}_{multi} (left bars) and \mathbf{p}_{mono} (right bars) for list types MEDHOM (dark bars) and HIGHHOM (lighter bars). Error bars are within-subject standard errors of the mean (Cousineau, 2005) (* = $p < 0.05$, ** = $p < 0.01$)

(mean P(YES) = 0.57) than on the MEDHOM list (mean P(YES) = 0.69), $t(9) = -4.99, p < 0.0001$. Furthermore, P(YES) for \mathbf{p}_{mono} on the HIGHHOM list (mean P(YES) = 0.52) was significantly lower than on the MEDHOM list (mean P(YES) = 0.57), although by a smaller amount, $t(9) = -3.29, p < 0.01$. Notably, all subjects but one showed both of these effects.

With increased list homogeneity, P(YES) values on both \mathbf{p}_{multi} and \mathbf{p}_{mono} were reduced. This rules out the possibility that the homogeneity effect arises solely from a change in the precision of representations in memory, as proposed by the *Memory-Precision hypothesis*. This hypothesis predicted that the P(YES) value for \mathbf{p}_{multi} would be lower on the HIGHHOM list as compared to the MEDHOM list (Equation 1), and that the P(YES) value for \mathbf{p}_{mono} would be equal or higher on the HIGHHOM list as compared to the MEDHOM list (Equation 2). The first of these two predictions is satisfied but the second is not. The data are, however, consistent with the predictions of the *Homogeneity-Computation hypothesis*, as described in Equations 3a and 3b.

Note that on both lists, the P(YES) values on the lure \mathbf{p}_{multi} are higher than the values for the target \mathbf{p}_{mono} . Rather than being an anomaly, this pattern is predicted by Global Matching Models. Since \mathbf{p}_{multi} is highly similar to multiple study items, the summation of these probe-item similarity values would be expected to produce a high P(YES) value. In contrast, even though \mathbf{p}_{mono} is a target having a high similarity value to the FAR study item, it is nonetheless dissimilar to the other study items; and hence the summation of probe-item similarity values would not produce any increased familiarity of the probe. This rationale was confirmed by evaluating the “outer” lure 1JND away from the FAR study item on the HIGHHOM study list. As this lure is a MONO-probe much like \mathbf{p}_{mono} , the associ-

ated P(YES) value should be less than that of \mathbf{p}_{mono} . This was indeed the case. The mean P(YES) value for this lure (0.38 ± 0.02) was significantly lower than the P(YES) value for the target probe \mathbf{p}_{mono} on this list (mean P(YES) = 0.52), $t(9) = -3.44, p < 0.01$.

Both *Memory-Precision* and *Homogeneity-Computation* hypotheses explicitly assume that the similarity of the probe to each of the study items plays no role in the origin of the homogeneity effect. If the probe’s similarity to the study items were indeed entirely responsible for the homogeneity effect then there should have been no difference in the P(YES) values for \mathbf{p}_{multi} and \mathbf{p}_{mono} between MEDHOM and HIGHHOM, as the similarity of these probes to each of the study items was equalized across MEDHOM and HIGHHOM. However, this prediction is clearly not true as shown in Figure 3. Therefore, our data demonstrate that the homogeneity effect cannot solely be due to the probe-item computations involved in evaluating the familiarity of the probe stimulus.

General Discussion

Our data provide evidence for the *Homogeneity-Computation* hypothesis (Kahana & Sekuler, 2002; Nosofsky & Kantner, 2006). Specifically, the data support the idea that a comparison process operates on the study items’ similarity, generating a signal that influences recognition judgments independent of the probe’s similarity to the study item.

We should note one potential confound in the design of our experiment: on MEDHOM study lists, \mathbf{p}_{multi} lies between the CLOSE and MIDDLE study items, but on HIGHHOM study lists, \mathbf{p}_{multi} lies outside these two items (as shown in Figure 1). Even though \mathbf{p}_{multi} has the same distance to the CLOSE, MIDDLE and FAR study items on both lists, it might be that this difference in its location could have influenced subject’s judgments, perhaps because of what have been called “edge effects” (Braidat et al., 1984). As the CLOSE study item along with the FAR study item define the boundaries (“edges”) of the interval within which the spatial frequencies of all the study items lie on each trial, it is possible that there there may have been differences in how subjects evaluate \mathbf{p}_{multi} on lists MEDHOM and HIGHHOM. However, such “edge effects” can be ruled out as an explanation as they do not account for the observed difference in P(YES) for the probe \mathbf{p}_{mono} between these two lists, as this probe matches the FAR study item that lies on the “edge” of both lists.

In conclusion, the *Homogeneity Computation* hypothesis suggests that recognition begins prior to the presentation of the probe, with the comparison of the study items to one another. We propose that the homogeneity signal is computed during the encoding of the study items. When the first item of the sequence, s_1 , is seen, it is represented and held in memory. When the second item in the sequence, s_2 , is presented, it is automatically compared to the memory representation of s_1 to produce a similarity signal, which is held in memory. When the third item, s_3 ,

is presented, it is automatically compared to the memory representations of s_1 and s_2 . The two resulting similarity signals are added to the first similarity signal, and the sum is scaled to produce a single value representing the study list's degree of "homogeneity." Note that with this process, the computation of the homogeneity signal would be completed with the presentation of the last study item. If this is indeed the case, then the generation of the homogeneity signal would not depend strongly on the length of the delay between the presentation of the last study item and the probe. However, such a process imposes an additional memory requirement, of maintaining the partially-computed homogeneity signal until the final study item is presented. This value may be stored in memory buffers related to the monitoring of trial difficulty. It seems remarkable that the homogeneity signal would be based on comparisons of this form. As the number of inter-item comparisons increases as a polynomial function of list length, assessing homogeneity could impose a greater computational burden than evaluating the familiarity of the probe.

References

- Bennett, P., & Cortese, F. (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research*, 36(2), 233–238.
- Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception. XIII. perceptual anchor model of context-coding. *Journal of the Acoustical Society of America*, 76(3), 722–731.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, 3, 37–60.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, 1(1), 42–45.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci*, 10(1), 14–23.
- Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Research*, 42(18), 2177–2192.
- Kahana, M. J., Zhou, E., Geller, A., & Sekuler, R. (2007). Lure-similarity affects episodic recognition: Detailed tests of a noisy exemplar model. *Memory & Cognition*, 35(6), 1222–1232.
- Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General*, 132(3), 351–78.
- Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neuroscience*, 23(6), 247–251.
- Magnussen, S., Greenlee, M., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, 31(7–8), 1213–1219.
- Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, 263(5146), 520–522.
- Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1), 3–27.
- Nosofsky, R. M., & Kantner, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition*, 34(1), 112–124.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317–1323.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850), 338–340.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153(736), 652–654.
- Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *PLoS Biology*, 5(3), e56.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.
- Yotsumoto, Y., Kahana, M. J., Wilson, H. R., & Sekuler, R. (2007). Recognition memory for realistic synthetic faces. *Memory & Cognition*, 35(6), 1233–1244.
- Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *J Exp Psychol Learn Mem Cogn*, 27(4), 1022–1041.
- Zhou, E., Kahana, M. J., & Sekuler, R. (2004). Short-term episodic memory for visual textures: a roving probe gathers some memory. *Psychological Science*, 15, 112–118.