

A stimulus-oriented approach to memory

Robert Sekuler
Volen Center for Complex Systems
Brandeis University

Michael J. Kahana
Department of Psychology
University of Pennsylvania

ABSTRACT – Our understanding of short-term recognition memory can be enhanced by careful choice and control of test materials. Theory-driven manipulation of memory test stimuli, including visual textures, human faces, and complex sounds, can reduce individual differences and make it possible to predict recognition performance for specific combinations of stimulus items. This stimulus-oriented approach to memory reveals that stimulus similarity plays two different important roles in recognition memory. By exploiting tools used in psychophysics, it is possible to generate mnemonic functions, “snapshots” that capture key features of variation in subjects’ memory strength.

KEYWORDS – recognition; short-term memory; models; criterion; decision making; memory

In his monumental *Handbook of Experimental Psychology*, S. S. Stevens (1951, p. 31) observed “In a sense there is only one problem of psychophysics, namely the definition of the stimulus. In this same sense there is only one problem in all of psychology –and it is the same problem.”

Careful attention to stimulus choice, design and control has led to many of sensory research’s key insights. We believe that a comparable, stimulus-oriented approach to memory affords equally significant theoretical leverage. And this belief stimulated our own collaboration, which draws on sensory research (R.S.) as well as memory research (M.J.K.). The work described in this article illustrates a few of the benefits that can come from a stimulus-oriented approach to memory. In particular, knowing the similarity relationships among memory test materials opens new ways to examine theoretical propositions about memory.

We do not claim to be the only people who appreciate the connections between memory and perception. In fact, others, too, are taken note of memory’s role in even the most basic perceptual tasks. This recent trend makes particularly good sense as some contemporary models of memory share assumptions with models of sensory discrimination. But, when memory models fail to link their stimulus representations to measures of perceptual similarity, they needlessly limit their ability to account for a variety of important phenomena.

Acknowledgments – Supported by NIH grants MH068404, MH55687 and MH61975. Many thanks to Kristina Visscher, Feng Zhou, Grace Hwang, Yuko Yotsumoto, Aaron Geller, Shivakumar Viswanathan and Marieke van Vugt for contributions to the research presented here. e-mail: sekuler@brandeis.edu or kahana@upenn.edu.

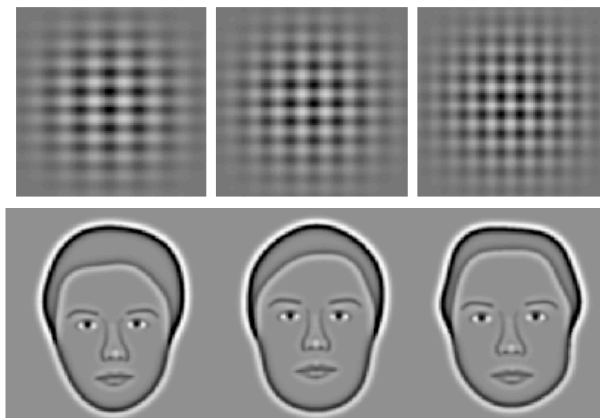


Figure 1. Examples of stimuli for our experiments. Top row: three compound gratings in which both horizontal and vertical spatial frequencies (number of bars) increase by about 15% between adjacent gratings. Bottom row: Sample synthetic faces used in stimulus-oriented study of memory. Each of these exemplars is derived from a different real face, and each lies the same metric distance away from a face that is the average of a large set of real faces.

To examine short term memory we exploit test stimuli that are little burdened by the complexities of extra-laboratory associations. Unlike semantically-rich items, such as words or images of objects that are easily categorized or named, our stimuli resist symbolic coding and mediation. These stimuli include visual textures (gratings, known also as Gabor patches). The top row of Figure 1 shows samples of these stimuli, which are constructed by combining sinusoidal gratings of differing orientations and/or spatial frequencies (number of bars per degree of visual angle). Controlled variation in test materials allows memory and

decision processes to be manipulated as some particular theoretical goal requires. This facilitates tests of theoretical claims about the ways in which memory's contents are expressed in recognition judgments. In our experiments, the similarity relations among stimuli are designed to challenge correct recognition and to produce errors, which are rich fodder for modeling. In addition, stimulus materials can be tailored to each subject's own sensory discrimination capacity. This limits sensory function's ability to exaggerate differences in memory function from one individual to another.

GLOBAL MATCHING MODELS

In one common recognition task, a series of study items is presented at the start of each trial. This series is followed, after a short retention interval, by a probe item. The subject's task is to decide whether this probe (\mathbf{p}) replicates any of the study items just seen. On some trials, called *target* trials, \mathbf{p} replicates one of the study items; on the remaining trials, called *lure* trials, \mathbf{p} replicates none of the study items.

Early models of recognition memory assumed that the receipt of the probe produced some signal of that stimulus' familiarity. This strength of signal could then be used to distinguish novel from repeated occurrences of an item, that is, to distinguish *lure* from *target* trials. This traditional account ran smack into two problems, one conceptual and one empirical. Conceptually, this account could not explain how, from all the memories that were stored on a trial, some one critical memory could be isolated and used in generating the familiarity signal. Empirically, the account could not explain why recognition judgments was governed not only by the probe's similarity to the lone matching study item, but by its similarity to other study items as well.

To overcome both these problems, global matching models were introduced. Such models assume that a recognition judgment reflects just a single value that summarizes information about the multiple items stored in memory. For example, in our own model, that single value represents "summed similarity", the sum of the \mathbf{p} 's similarity to each and every one of the study items.

Figure 2A is a diagram of the global matching component of our own model. We assume that each study item generates its own representation or exemplar in memory. Following the tradition of multidimensional signal-detection theory (Ashby & Maddox, 1998), any one stimulus' representation varies from occasion to occasion. This variation is represented in Figure 2 by the "clouds" that correspond to each exemplar.

To generate the value of summed similarity, \mathbf{p} 's similarity to each study item is assessed and those separate assessments are summed. On average, summed similarity values produced on *target* trials will exceed ones produced on *lure* trials (because every *target* trial, but not every *lure* trial, will have at least one exem-

plar that is highly similar to \mathbf{p}). Incidentally, lest this summed similarity computation seem too fanciful, note that researchers have identified neural circuits capable of carrying out this computation, e.g., familiarity-sensitive neurons found in the perirhinal cortex (Bogacz & Brown, 2003), a region in the medial temporal lobe.

The diagram in Figure 2A shows that once summed similarity has been computed, its value sent to a decision module where summed similarity is compared against a criterion. The criterion value reflects (i) the *a priori* probability that the probe would have replicated a study item, and (ii) the cost of a false recognition, along with the value of a correct one. If summed similarity exceeds the criterion, the probe is judged "old" (it replicates one of the study items); otherwise, the probe is judged "new".

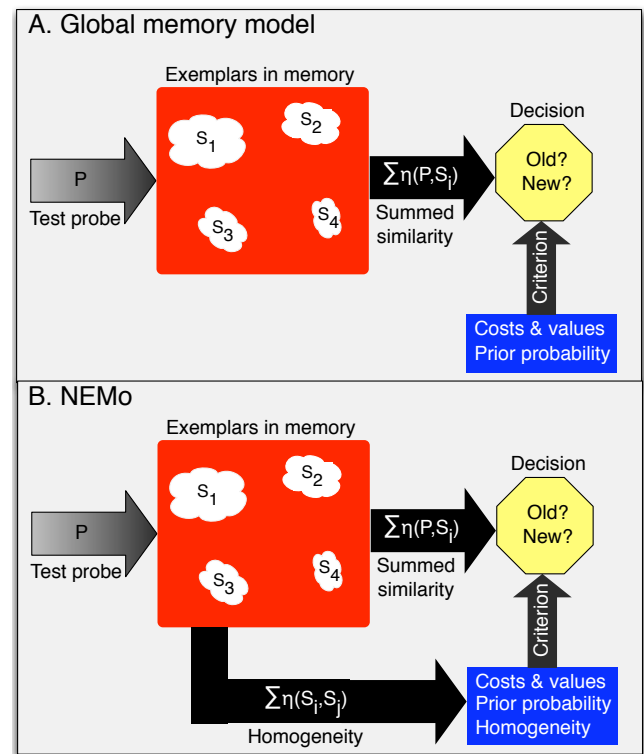


Figure 2. Schematic diagram illustrating basic components of global matching models (Panel A) and NEMO, the Noisy Exemplar Model (Panel B). In both models, the presentation of four study items, $s_1 \dots s_4$, gives rise to exemplars of those items in memory. The text describes the operations that the models perform on these exemplars in order to decide whether the test probe is "old" or "new".

FINDING NEMO: STUDY ITEM HOMOGENEITY

In our first experiments, using gratings as test materials (Kahana & Sekuler, 2002), the summed similarity model shown in Figure 2A produced a passable, but obviously imperfect account of the results. For some

trials in particular, the model's predictions were way off; something was clearly wrong. To find out what, we examined our results through the lens of the similarity relations among each trial's stimuli. Knowing what the similarity relations among stimuli were, we characterized trials on which recognition performance departed significantly from predictions based on summed similarity alone (Kahana & Sekuler, 2002). A close examination of the "deviant" trials revealed the operation of a previously-hidden, but powerful influence on recognition performance. And from this discovery, our Noisy Exemplar Model (NEMo) was born.

Reflecting the analysis of deviant trials in our first experiments, Figure 2B highlights NEMo's key new element, an additional influence on the criterion value: the homogeneity of the study items (Nosofsky & Kantner, 2006). Functionally, the impact of summed similarity is modulated by study item homogeneity. Empirically, when study items are similar to one another, the criterion value is increased, reducing the likelihood of false recognitions (saying "yes" on *lure* trials). So with greater homogeneity, subjects demonstrate increased ability not to be fooled into falsely recognizing probes that resemble but do not perfectly match one of the study items. Note that what seems to be an advantage—being fooled less often—comes at a price, which is reduced likelihood of correct recognitions as well.

It is important to note that the impact of study-item homogeneity is not restricted to recognition of Gabor patches, but has been confirmed with various other stimuli, as diverse as color patches (Nosofsky & Kantner, 2006), realistic, synthetic human faces (Yotsumoto, Kahana, Wilson, & Sekuler, in press), and abstract, complex sounds (Visscher, Kaplan, Kahana, & Sekuler, 2007). Note that without some *a priori*, stimulus-oriented characterization of study-item homogeneity, this powerful influence on recognition performance would have gone undetected, lost in the data's unexplained variance. However, by augmenting the basic, summed-similarity framework with the idea that within-list summed similarity influences recognition decisions, NEMo fulfills one goal of any model: shifting substantial variance in the data from the "random" (unexplained) column to the "deterministic" (accounted for) column.

MEMORIES COME, MEMORIES GO (QUICKLY)

A stimulus-oriented approach to memory also makes it possible to examine the temporal ebb and flow of elements that influence recognition. For example, Visscher et al. (2007) showed that in contrast to other determinants of the subject's criterion, which develop over many trials, homogeneity's entire influence develops quickly, within just single trial, and then dissipates just as quickly. Visscher et al. (2007) measured recognition memory for so-called moving ripple stimuli, broadband sounds that vary sinusoidally in both

time and in frequency content. These stimuli are generated by superimposing sounds at many frequencies whose loudness vary over time, have a formal resemblance to most speech sounds.

Visscher et al. (2007) constructed two types of *lure* trials, ones in which the study items were homogeneous (the study items were only one just noticeable difference apart) and ones in which the study items were heterogeneous (the study items were at least four just noticeable differences apart). The former were expected to produce a high criterion value for summed similarity, the latter were expected to produce a low criterion value for summed similarity. The two types of *lure* trials were intermixed among various other trials. The mixture of trials was adjusted so that identical, critical *L* trials followed either 'low criterion' trials (heterogeneous study items) or 'high criterion' trials (homogeneous study items).

This portion of Visscher et al. (2007)'s larger study produced two noteworthy results related to false recognitions. First, as NEMo predicts, low-criterion trials produced far higher false alarm rates than high criterion trials did. Second, on the immediately succeeding, neutral trials, there was no evidence of the criterion differences that had been generated on the previous trial: false alarm rates on trials that *followed* high criterion trials were no different from false alarm rates on trials that *followed* low criterion trials. So, homogeneity's influence appears to reach full strength during the course of a single trial, as subjects respond to the study items on that trial, and then recedes immediately thereafter. Note that this remarkable encapsulation of criterion change stands in contrast to the inter-trial, proactive interference observed in other settings, with semantically-related stimuli.

MNEMOMETRIC FUNCTIONS: SNAPSHOTS OF MEMORY

For over a century, psychophysical research has exploited a tool called the psychometric function. This function relates the likelihood of some psychophysical judgement, such as "yes, I see it", to a measure of stimulus strength, such as light level. Early in the last century, a few researchers advocated that an analogous function would be valuable for memory research (Williams, Titchener, & Boring, 1918; Ernst, Smith, Moessner, Rudisill, & Atwater, 1924). Recently, Zhou, Kahana, and Sekuler (2004) introduced such a function, which they called a mnemometric function.

To understand what the mnemometric function is, imagine that some stimulus is presented and remembered. If the subject's memory is probed with that exact same stimulus, there is high probability that it will be recognized. So over many trials, the proportion of recognition responses, $P(\text{yes})$, to that probe will be high. However, if memory is probed with a stimulus that is very dissimilar to the remembered stimu-

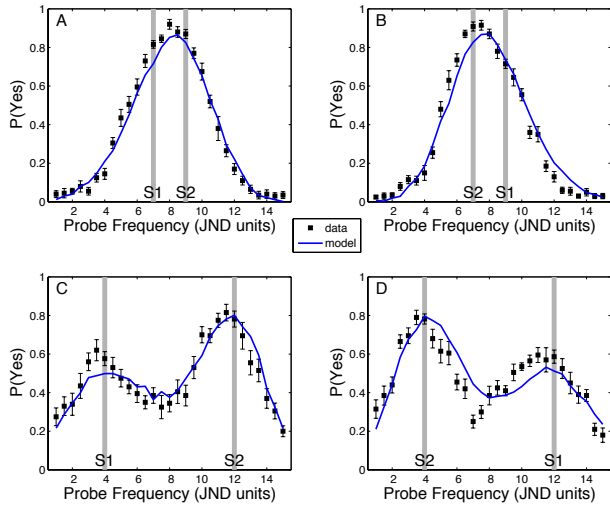


Figure 3. Mnemonic functions generated as p varied in spatial frequency relative to the frequencies of two study items. Study item frequencies are shown by the vertical gray lines labeled S1 and S2, signifying the first and second study item, respectively. Panels A and B: mnemonic functions with study items that differed by just two JNDs; Panels C and D: mnemonic functions with study items that differed by eight JNDs. Panels A and C are for trials on which the second study item's spatial frequency was lower than the first study item's; Panels B and D are for the reverse case. Each mean value of $P(\text{Yes})$ is bracketed by ± 1 standard error. The solid blue line in each panel represents the values predicted by NEMO. Adapted from Kahana et al. (in press)

lus, $P(\text{yes})$ will be low. Between these two extreme cases there are intermediate cases: probes of varying levels of similarity to the original stimulus. And these probes should evoke varying, intermediate values of $P(\text{yes})$.

The mnemonic function, examples of which are shown in Figure 3, is a kind of snapshot of memory strength. It relates a subject's recognition responses to the similarity between a probe and one or more study items. As the probe varies (or "roves") along some perceptual dimension, the responses it evokes on multiple trials trace out variation in memory strength. The mnemonic functions shown in Figure 3 come from a study (Kahana et al., in press) in which two study item gratings were presented on each trial. Although the items' spatial frequencies varied from trial to trial, the difference between study items was constrained: sometimes the items were separated by two just noticeable differences (Panels A and B), other times they were separated by eight just noticeable differences.

The mnemonic functions shown in Figure 3 demonstrate several important features of recognition memory. First, the clear bimodality seen in Panels C and D demonstrates that the study items are represented in memory as exemplars rather than as a simple aggregate prototype. Second, the spread of responses around each mode indicates the variability associated

with the exemplars. This variability makes it impossible to discern bimodality in Figure 3A and B, where study items were very similar to one another. Third, the relative advantage in memory enjoyed by the study item that is presented last (the recency effect) produces near-mirror image relationships between the functions in Panels C and D. The higher and lower modes seem to swap locations between the panels because the relative spatial frequencies of the first and second study items was swapped. On some trials (Panel C), the second study item had higher spatial frequency than the first study item; on other trials (Panel D) the reverse was true. In both cases, the mode associated with the second study item is higher than with the first study item. Fourth, these mnemonic functions and others as well, show no evidence of systematic shift in average memory away from the actual spatial frequencies of the study items. Finally, the blue lines in each panel show predictions made by NEMO. When NEMO was compared to various alternative models, Kahana et al. (in press) found that NEMO provided a far better fit to the mnemonic functions than did either standard global matching models, or models in which recognition decisions are based only on the single study item most similar to the probe.

FUTURE DIRECTIONS

Despite the successes of our stimulus-oriented approach to memory, much work lies ahead. Whereas current modeling efforts have been limited to fitting data from individual lists, the human brain is able to focus retrieval on one target list while still some retaining information learned on previous lists. In other words, memory's slate may not be wiped entirely clean after each trial. To address this, models will need to incorporate a mechanism for coding memories within their temporal context and using contextual information to target specific memories. Also, we expect that mnemonic functions will find additional uses, both theoretical and applied. For example, because they afford rich information about similarity-based confusions in recognition memory, mnemonic functions could be a valuable tool for a quantitative assessment of eyewitness testimony, including lineup identification. Another exciting challenge is to identify the neural circuits that compute summed similarity. This effort could start by correlating brain signals measured with non-invasive EEG, MEG, or fMRI methods, with model-derived values of summed similarity and homogeneity on each trial of an experiment. Finally, advances in computational methods will enable us to dimensionalize and compute similarity relations for complex natural stimuli. For example, new computational techniques, operating on large digital databases of potential stimuli, could enable researchers to apply similarity-based models to verbal materials. These and other developments promise to extend far beyond the lab-

oratory the power of a stimulus-oriented approach to memory.

Recommended Reading

A clearly written, user-friendly introduction for readers who wish to expand their knowledge of basic issues and approaches to classification and related cognitive functions such as recognition: Estes, W. K. (1994) *Classification and cognition*. Oxford: Oxford University Press.

A representative application of a stimulus-oriented approach to memory, including comprehensive tests of alternative accounts: Kahana, Zhou, Geller & Sekuler (in press). (See references)

A review and comparison of memory models, including global matching models: 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.

A review of findings on memory for sensory stimuli and of the contribution of various brain regions to such memory: Pasternak, T. & Greenlee, M. W. (2005) Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6, 97-107.

A description of how a stimulus-oriented approach can be applied to age-related changes in memory: Sekuler, R., McLaughlin, C., Kahana, M. J., Wingfield, A. & Yotsumoto, Y (2006) Short-term visual recognition and temporal order memory are both well-preserved in aging. *Psychology & Aging*, 21, 632-637.

References

- Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (p. 251-301). Mahwah, NJ: Lawrence Erlbaum and Associates.
- Bogacz, R., & Brown, M. W. (2003). Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus*, 13(4), 494-524.
- Ernst, J. L., Smith, F. E., Moessner, L. R., Rudisill, E. S., & Atwater, M. J. (1924). Further data for an associative limen. *American Journal of Psychology*, 35, 255-261.
- Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Research*, 42(18), 2177-192.
- Kahana, M. J., Zhou, F. A., Geller, A., & Sekuler, R. (in press). Lure-similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model. *Memory & Cognition*.
- Nosofsky, R. M., & Kantner, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition*, 34(1), 112-24.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (p. 1-49). New York: J Wiley & Sons.
- Visscher, K., Kaplan, K., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *Public Library of Science, Biology*, 5, e56.
- Williams, H. D., Titchener, E. B., & Boring, E. G. (1918). Minor studies from the psychological laboratory of Cornell University: XL. On the calculation of an associative limen. *American Journal of Psychology*, 29, 219-226.
- Yotsumoto, Y., Kahana, M. J., Wilson, H. R., & Sekuler, R. (in press). Recognition memory for realistic synthetic faces. *Memory & Cognition*.
- Zhou, F., Kahana, M. J., & Sekuler, R. (2004). Short-term episodic memory for visual textures: a roving probe gathers some memory. *Psychological Science*, 15(2), 112-8.