

Lure similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model

MICHAEL J. KAHANA

University of Pennsylvania, Philadelphia, Pennsylvania

FENG ZHOU

Brandeis University, Waltham, Massachusetts

AARON S. GELLER

University of Pennsylvania, Philadelphia, Pennsylvania

AND

ROBERT SEKULER

Brandeis University, Waltham, Massachusetts

Summed-similarity models of visual episodic recognition memory successfully predict the variation in false alarm rates across different test items. With data averaged across subjects, Kahana and Sekuler (2002) demonstrated that subjects' performance appears to change along with the mean similarity among study items; with high interstimulus similarity, subjects were less likely to commit false alarms to similar lures. We examined this effect in detail by systematically varying the coordinates of study and test items along a critical stimulus dimension and measuring memory performance at each point. To reduce uncontrolled variance associated with individual differences in vision, the coordinates of study and test items were scaled according to each subject's discrimination threshold. Fitting each of four summed-similarity models to the individual subjects' data demonstrated a clear superiority for models that take account of interitem similarity on a trialwise basis.

Kahana and Sekuler (2002) adapted Sternberg's (1966, 1975) procedure to study episodic recognition memory for series of textures, which were created by linearly summing sinusoidal gratings. This adaptation made it possible to quantify and characterize interference in memory among successively presented stimuli. Unlike semantically rich stimuli, such as words or images of recognizable and nameable objects, multidimensional textures are not burdened by the complexities of extralaboratory associations, and they resist symbolic coding (Della-Maggiore et al., 2000; Hwang et al., 2005). Because of their well-defined, natural metric representations in a low-dimensional space (Kahana & Bennett, 1994), compound grating stimuli facilitate manipulation of interitem similarity relations, which are important determinants of visual episodic recognition (Kahana & Sekuler, 2002; Sekuler, Kahana, McLaughlin, Golomb, & Wingfield, 2005; Zhou, Kahana, & Sekuler, 2004). The availability of a natural stimulus metric for defining similarity relations among items enables detailed mathematical accounts of recognition memory to be applied to results from individual stimulus lists.

Using Nosofsky's (1984, 1986) generalized context model (GCM) as our starting point, we developed NEMO, a noisy exemplar model, which combines core aspects of

GCM with significant new assumptions. First, NEMO follows the tradition of multidimensional signal detection theory (e.g., Ashby & Maddox, 1998) in assuming that stimulus representations are coded in a noisy manner, with a different level of noise associated with each dimension. Second, NEMO augments the summed-similarity framework of item recognition (Brockdorff & Lamberts, 2000; Clark & Gronlund, 1996; Humphreys, Pike, Bain, & Tehan, 1989; Lamberts, Brockdorff, & Heit, 2003; Nosofsky, 1991, 1992) with the idea that within-list summed similarity (not just probe-to-list-item similarity) influences recognition decisions. In fitting NEMO to data from two experiments, Kahana and Sekuler (2002) found that subjects were more likely to say *yes* to lures following study of lists whose items had low interitem similarity than to lures following lists whose items had high interitem similarity. Subjects appear to interpret probe-to-list similarity in light of within-list similarity, with greater list homogeneity leading to a greater tendency to reject lures that are similar to one or more of the studied items. The impact of within-list similarity was confirmed by Nosofsky and Kantner (2006), using color patches as stimuli.

In vision research, many studies focus on the individual performance of a small number of subjects. In contrast,

M. J. Kahana, kahana@psych.upenn.edu

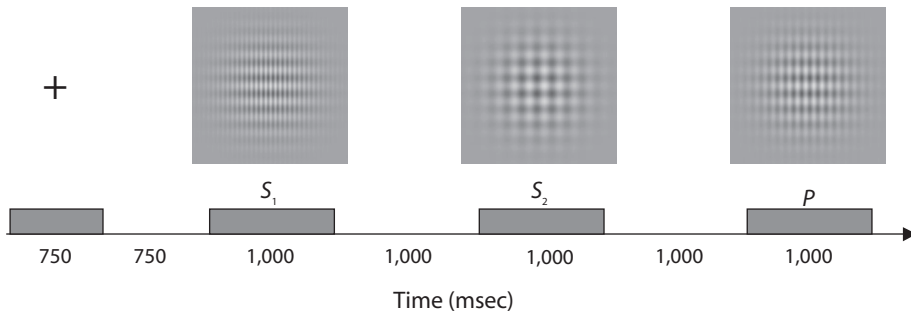


Figure 1. A schematized example of a trial. The events, whose durations are indicated on the timeline at the bottom of the figure, begin with a fixation point, which is then succeeded by s_1 , s_2 , and p . In this example, p is a lure, differing from both s_1 and s_2 . The correct response, therefore, would be *no*.

memory research tends to focus on performance averaged across subjects. The statistical advantage of averaging is obvious, but this advantage exacts a toll. It can introduce qualitative changes into the pattern of data, thereby distorting the outcome of quantitative modeling (e.g., Maddox, 1999). Although we subscribe to the goal of understanding the visual episodic recognition performance of individual subjects, previous data sets, including our own, were not large enough to support meaningful modeling of individual subjects' data. Therefore, our model-driven analyses to date have all been done on averaged data. The experiment described below was designed to produce data sufficient to support modeling on an individual-subject level.

Our goal in this article is to compare the fits of a suite of NEMO variants to individual subjects' data, where the latter are generated using the roving probe paradigm (Zhou et al., 2004). In the roving probe paradigm, subjects study a short list of stimuli that vary along one (or more) dimensions for a subsequent recognition memory test. By randomly selecting test probes from a uniform distribution along the same stimulus dimension as the studied items, the roving probe method can reveal the parametric relation between the probability of subjects' *yes* responses and the metric properties of the probe stimulus. We first will present the experimental study that generated a data set suitable for model fits to individual subjects' results. We then will describe the NEMO model in detail, including minor changes from our earlier implementation (Kahana & Sekuler, 2002). The third section of the article will present fits of four variants of NEMO to individual subjects' data and will discuss the contributions of the various core modeling assumptions to obtaining a good quantitative fit. Finally, we will discuss the limitations of NEMO's current version and ways that the model could be extended to account for a broader range of important attributes of perceptual memory.

Method

Subjects. The subjects were 5 male and 5 female volunteers whose ages ranged from 19 to 30 years. They had normal or corrected-to-normal visual acuity, as measured with Snellen targets, and normal contrast sensitivity, as measured with Pelli-Robson charts (Pelli, Robson, & Wilkins, 1988).

Apparatus. Stimuli were generated and displayed using MATLAB 5 and extensions from the Psychophysics and Video Toolboxes (Brainard, 1997; Pelli, 1997). The stimuli were presented on a 14-in. CRT computer 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. Mean screen luminance was maintained at 36 cd/m^2 .

Stimuli. On each trial, three compound gratings, s_1 , s_2 , and p , were presented sequentially. At the start of each trial, a fixation point was centered on the screen for 750 msec. The fixation point was followed by a 750-msec period of uniform luminance. This was followed by s_1 and then s_2 , each for 1,000 msec, separated by an interstimulus interval (ISI) of 1,000 msec. During the ISI, the display area was filled with uniform luminance (36 cd/m^2). Finally, 1,000 msec after s_2 , p was presented and remained visible until the observer responded. p disappeared 1,000 msec after its onset if the observer did not respond in such a period. A sample trial is schematized in Figure 1.

In each stimulus, one vertical and one horizontal sinusoidal luminance grating were superimposed, which generated a luminance profile, $L_{x,y}$, given by

$$L_{x,y} = L_{\text{avg}} \left[1 + \frac{1}{2} A_1 \cdot \cos(\pi f x) + \frac{1}{2} A_2 \cdot \cos(\pi g y) \right], \quad (1)$$

where L_{avg} is the mean luminance, f is the spatial frequency of the stimulus' vertical component (*vertical frequency*) in cycles per degree, g is the frequency of the horizontal component (*horizontal frequency*), and A_1 and A_2 , the Michelson contrasts for the two components, were set to 0.4, a value well above the threshold for detection. Each grating subtended 6° of visual angle at a viewing distance of 82 cm. To minimize edges, the stimuli were windowed by a circular 2-D Gaussian function with a space constant of 1° of visual angle.

Prior to memory testing, each subject's spatial frequency discrimination was measured with an up-down-transformed response procedure (Wetherill & Levitt, 1965). On each trial, a subject viewed two gratings, each presented for 750 msec, with a 1,000-msec ISI (note that this was the same timing as would be used in the memory experiment). The spatial frequency of one grating on each trial was drawn randomly from the set 1, 2, 3, 4, and 5 cycles/deg, which spanned the range of frequencies that would be used in the memory experiment. With equal probability, the second grating on each trial was either fractionally higher or lower in spatial frequency than the first. The subject judged which of the two gratings had higher spatial frequency, judging only vertical frequency.

This adaptive psychophysical procedure yielded the spatial frequency difference that produced 70.7% correct discrimination between gratings. The subjects' Weber fractions ranged from 0.073 to 0.189. Each subject's own Weber fraction was used to standardize the stimuli with which that subject's recognition memory was measured. The spatial frequency of the stimuli can be expressed in just notice-

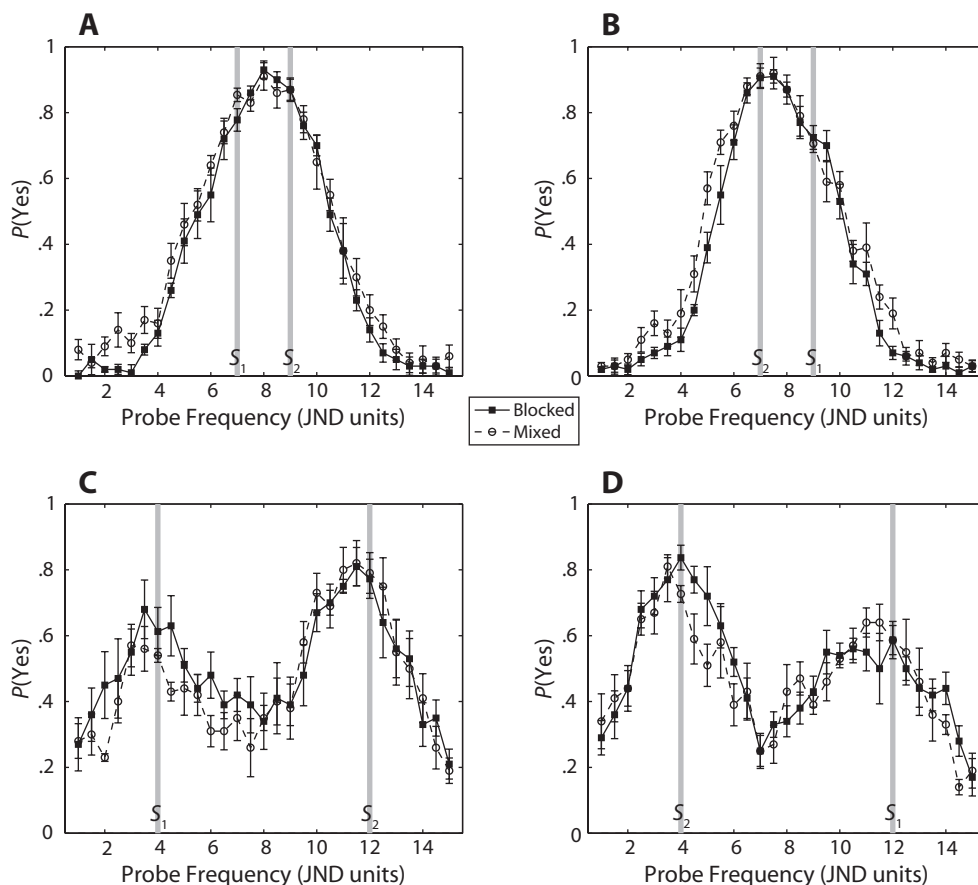


Figure 2. Mnemonic functions: Probabilities of *yes* responses as *proved* along the spatial frequency dimension. Solid lines indicate the blocked group data, and dashed lines indicate the mixed group data. The minimum spatial frequency of p is normalized to 1. The normalized spatial frequencies of s_1 and s_2 are indicated by gray bars. The upper panels show data from the near conditions ($|s_1 - s_2| = 2$ just noticeable difference [JND] units); the lower panels show data for the far conditions ($|s_1 - s_2| = 8$ JND units). The panels on the left show data from the conditions in which s_1 had the lower spatial frequency of the two study items; the panels on the right show data from the conditions in which s_1 had the higher spatial frequency. Error bars represent ± 1 standard error of the mean.

able difference (JND) units, which is the product of spatial frequency and the individual's Weber fraction. This stimulus standardization, which was introduced by Zhou et al. (2004), minimized visual encoding as a source of individual differences in performance.

On each trial, the same spatial frequency was used for the horizontal component of all three gratings, s_1 , s_2 , and p ; differences among the items were generated only by variation in their vertical frequencies. Between trials, the horizontal frequency varied randomly between 0 (no luminance variation along the horizontal dimension) and 3 cycles/deg. The vertical frequency is determined by both an individual subject's JND and interstimulus configuration. The geometric mean spatial frequency of the two study gratings was 7 JNDs above a minimum reference value ranging from 1 to 2 cycles/deg, which was labeled as 1 on the abscissa in Figure 2. The difference between the study gratings' vertical frequencies, $|s_1 - s_2|$, was either 2 or 8 JNDs. These two values occurred randomly but equally often. On half the trials, s_1 's vertical spatial frequency was lower than s_2 's; on the remaining trials, the reverse was true. In 29 steps of half a JND, p 's frequency ranged from 0 to 14 JNDs above the lowest reference value. As is customary in Sternberg-type experiments, on half the trials, p matched one of the two study items (we designate such ps targets). On the remaining trials, p matched neither of the study items (we designate such ps lures). Because a target probe was

equally likely to replicate either one of the study items, the observers had to attend to both.

To minimize the subjects' ability to base judgments exclusively on local, retinotopic matches, the absolute phases of horizontal and vertical components were varied randomly from stimulus to stimulus within each trial.

To determine whether the subjects' performance with variable interitem similarity differed from that with constant interitem similarity, for 5 subjects, $|s_1 - s_2|$ values were randomly mixed in a session; these subjects comprised the *mixed* group. For the remaining subjects, in the *blocked* group, $|s_1 - s_2|$ values were held constant during blocks of trials within a session.

Procedure. The subjects served in 10 sessions, which were separated by 24–72 h. During testing, the subjects sat with head supported by a chin-and-forehead rest, viewing the computer display binocularly from a distance of 82 cm. A trial was initiated by the press of a key on the computer keyboard. The subjects were instructed to respond as accurately and quickly as possible. By pressing computer keys representing *yes* and *no*, the subjects signaled their judgment as to whether p was identical to one of the study items, either s_1 or s_2 , or different from both study items. The computer generated brief, distinctly different tones after correct and incorrect responses, providing the subjects with trialwise knowledge of the results. Throughout

each session, a webcam broadcast a video feed of the subject's face over the laboratory's local area network, allowing the experimenter to monitor subject compliance.

Results

Zhou et al. (2004) introduced the term *mnemometric function* to describe the relation, in a recognition experiment, between the proportion of *yes* responses and the metric properties of a probe stimulus. In their experiment, as the probe item *roved* or varied in spatial frequency, the roving probe sampled memory strength at various points along the spatial frequency continuum, sweeping out a probability function that afforded a snapshot of the distribution of memory strength.

Figure 2 shows the average mnemometric functions for the four different conditions in the present experiment. The upper panels show data from the conditions in which the difference between the study gratings' vertical frequencies $|s_1 - s_2|$ was 2 JNDs (the *near* condition); the lower panels show data from the conditions in which the difference was 8 JNDs (the *far* condition). The panels on the left show data from the conditions in which s_1 had a lower spatial frequency; the panels on the right show data from the conditions in which s_1 had the higher spatial frequency.

As was revealed by multiple *t* tests, only 5 of 116 data points on the mnemometric functions differ significantly between the mixed and the blocked groups. In two cases, $P(\text{yes})$ of the mixed group is significantly higher than that of the blocked group ($p < .05$ without family-wise error adjustment). The two groups performed similarly overall. The small, unsystematic differences between groups will not be considered further.

As is shown in all four panels, false alarms increased as the spatial frequency of a lure approached that of one of the two studied items. When the spatial frequencies of the studied items were separated by just 2 JNDs, the highest false alarm rates were observed when the lure's frequency lay between those of the two studied items. When the studied items were separated by 8 JNDs, however, the false alarm rate decreased as the lure's spatial frequency deviated from that of either of the two studied items. False alarms also exhibited a recency effect insofar as they were greater when the lure was similar to the more recently studied test item.

NEMO: A Noisy Exemplar Model

Building on exemplar models of classification and recognition (McKinley & Nosofsky, 1996; Nosofsky, 1986), we assume that as each stimulus is presented, its feature space coordinates are stored in memory and that judgments are based largely on the summed similarity between the probe item and these stored representations. In particular, summed similarity refers to the sum of pairwise similarity measures between the probe, on one hand, and the representations of each of the study items, on the other. Borrowing ideas from decision-bound models of human classification (Ashby & Maddox, 1998; Ennis, Palen, & Mullen, 1988; Maddox & Ashby, 1996), NEMO represents each stimulus as a multivariate normal distribution in feature space and uses a deterministic response rule, re-

sponding *yes* if the summed similarity crosses a decision bound (criterion) that separates targets and lures.

The basic computation underlying recognition performance is the summed pairwise similarity between each item's noisy representation and the relatively noiseless representation of the probe. If stimuli are randomly selected from a multidimensional space, the summed similarity of a target to the contents of memory will typically exceed the summed similarity for a lure. This provides a basis for modeling two-alternative forced choice recognition. However, for *yes-no* recognition, as in the Sternberg (1966) procedure used here, the summed similarity value cannot be used directly but must be compared with some experience-based threshold value that distinguishes between targets and lures.

We assume that a subject uses an optimal decision criterion to decide whether the summed similarity is more likely to have come from the presentation of a lure or from the reappearance, as a probe item, of a stimulus that was in the list. Presumably, experience enables subjects to adjust their criteria so as to suit the context (e.g., Petzold & Haubensak, 2004; Treisman & Williams, 1984). Morgan, Watamaniuk, and McKee (2000) provided a striking demonstration of the ease and flexibility with which such adjustments in criterion can be made from one trial to the next. Their demonstration is supported by the present study's finding of small, nonsystematic differences in performance between subjects who made up the mixed and the blocked groups.

As in Nosofsky (1986), we define the similarity, η_{s_i, s_j} , among two representations, s_i and s_j , as given by

$$\eta(s_i, s_j) = e^{-\tau d(s_i, s_j)}, \quad (2)$$

where d is the weighted distance between the two stimulus vectors and τ determines the steepness of the exponential generalization gradient (Chater & Vitányi, 2003; Shepard, 1987). Increasing the value of τ causes similarity to decrease more rapidly with increasing distance. The distance along each dimension $|s_i(k) - s_j(k)|$ is weighted by a factor w_k , to ensure (1) that the measurement is not sensitive to absolute variations in the scale of the dimensions and (2) that the model can capture global differences in the attention that each dimension attracts. In fitting data from the present experiment, in which stimuli varied along just one dimension, we set the weighting parameter to 1.

We assume that each stimulus is stored imperfectly in memory. To the stored representation of each item, we add a noise vector, ϵ , whose components are zero-mean Gaussian variables whose variance depends on the stimulus dimensions comprising that item and on the recency of that item's occurrence. Variability in subjects' responses from one occurrence of an item to another are modeled by the sampling of each item from its noisy representation. We use two mechanisms to simulate forgetting: (1) We assume that the most recent stimulus contributes the most to the summed similarity and that earlier items contribute less (the α parameter in Equation 3, below), and (2) we assume that the stimulus coordinates of older representations are coded with greater degrees of noise (i.e., larger variance of ϵ).

NEMO's most important innovation is the introduction of an interstimulus similarity term into the summed-similarity computation (see Kahana & Sekuler, 2002, for a detailed discussion of the importance of this parameter). Given a list of items, $s_1 \dots s_L$, and a probe item, p , NEMO will respond *yes* if

$$\underbrace{\sum_{i=1}^L \alpha_i \eta(p, s_i + \varepsilon_i)}_{\text{Summed probe item similarity}} + \underbrace{\frac{2}{L(L-1)} \beta \sum_{i=1}^{L-1} \sum_{j=i+1}^L \eta(s_i + \varepsilon_i, s_j + \varepsilon_j)}_{\text{Mean interitem similarity}}, \quad (3)$$

exceeds a threshold value, C . If $\beta = 0$, the model reduces to a close variant of Nosofsky's GCM model. If $\beta < 0$, a given lure will be more tempting when s_1 and s_2 are widely separated; conversely, if $\beta > 0$, a lure will be less tempting when s_1 and s_2 are widely separated.

The α and ε terms in Equation 3 determine the rate of forgetting. For the most recent item, $\alpha = 1$, but for earlier items, α is allowed to take on smaller values. In simulating memory for a list comprising just two stimuli, a single α parameter would determine the contribution of the older item to the summed-similarity calculation. Forgetting is also modeled by allowing the variance-covariance matrix of ε to change as a function of stimulus recency. Kahana and Sekuler (2002) fit NEMO to data from lists of four items that varied along three relevant dimensions. The number of items and critical dimensions meant that a full estimation of the variance-covariance matrices for each serial position would have required estimating 24 parameters. Kahana and Sekuler therefore assumed that the covariances were zero and that the noise was fixed across serial position. In the present application to stimuli that vary along just a single critical dimension, we consider the consequences of allowing the variance to change across serial position. This entails estimation of separate variance parameters for each serial position.

In the next section, we will consider the consequences of replacing the exponential similarity distance function with a binary function in which similarity falls from 1 to 0 when the distance exceeds a critical threshold, Δ (this relation is characterized by a Heaviside function). Our motivation for considering such a binary relation between similarity and distance comes from Zhou et al.'s (2004) report that a signal detection model in which only the nearest item contributes to the recognition decision provided a reasonably good qualitative account of the mnemonic functions obtained using the roving probe technique.¹ In comparing these two variants of NEMO, we hope to determine whether an exponential or a binary summed-similarity model provides a better account of observers' visual recognition memory performance.

Fitting NEMO to Mnemonic Functions

We fit four different versions of NEMO to each subject's mnemonic functions by finding the parameters' values that minimized the root-mean-squared differ-

Table 1
Factorial Arrangement of the Four Variants of NEMO Fit to Experimental Data

Interstimulus Similarity	Similarity-Distance Function	
	Exponential	Heaviside
Yes	NEMO _{βE}	NEMO _{βH}
No	NEMO _E	NEMO _H

Note—In model designations, the subscripts E and H denote exponential and Heaviside similarity-distance functions, respectively; the subscript β denotes the inclusion of an interitem similarity term.

ence (RMSD) between observed and predicted values.² The four versions represent factorial combinations of (1) whether interstimulus similarity was or was not allowed to influence the response and (2) whether similarity fell exponentially with physical distance or whether similarity was forced to be binary, falling from 1 to 0, when the probe item distance exceeded some critical threshold. (As was mentioned earlier, this binary relation is characterized by a Heaviside function.) We label these four variants NEMO _{β E}, NEMO _{β H}, NEMO_E, and NEMO_H, as defined in Table 1.

Because they possess an additional free parameter (β), NEMO _{β E} and NEMO _{β H} have an inherent advantage over NEMO_E and NEMO_H. To compare all four models on a more equal footing, we computed Schwarz's (1978) Bayesian information criterion (BIC) for each model and subject. Under the simplifying assumption of normally distributed error terms, BIC is simply

$$\text{BIC} = k \ln(n) + n \ln \left(\frac{\text{RSS}}{n} \right), \quad (4)$$

where k is the number of free parameters, n is the number of observations, and RSS is the residual sum of squares. Better-fitting models will have lower BIC values.

The left column of Figure 3 shows the difference between observed and predicted values for the fits of NEMO _{β E} and NEMO_E to the average subject data; the middle and right columns show fits to two representative subjects. Later, we will present formal analyses of each model's fit to each of the 10 individual subjects. Figure 4 shows the differences between observed and predicted values for the fits of NEMO _{β H} and NEMO_H.

Inspection of these graphs shows that NEMO _{β E} (exponential, interstimulus similarity) provided the best fit to the average data (RMSD = 0.070, BIC = -635). The other models generated RMSDs ranging from 0.087 to 0.089 and BIC values ranging from -534 to -551. Whereas the inclusion of interstimulus similarity substantially improved the fit of the exponential summed-similarity models, it had little effect on the fit of the Heaviside models (see Figure 4).

Figures 3 and 4 also reveal systematic deviations between the model and the data for all four NEMO variants. In the near condition, the probability of making a *yes* response drops too slowly as the probe moves away from s_1 and s_2 ; this is observed in the peaks around JND = 4 and JND = 12 in panels A and B. In the far condition, there appears to be some systematic bias in the data that is not

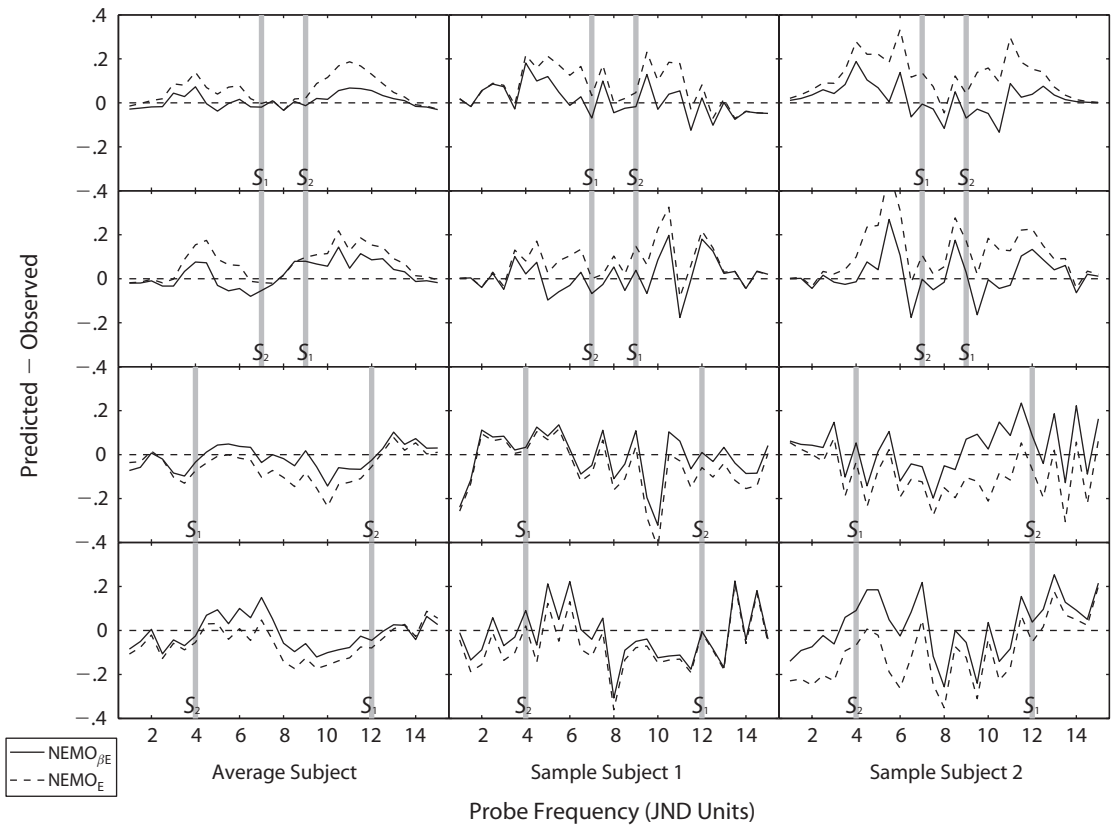


Figure 3. Difference between predicted and observed mnemometric functions for exponential variants of NEMO with and without an adjustment for interstimulus similarity (i.e., $NEMO_{\beta E}$, solid lines; $NEMO_E$, dashed lines). Residuals in the left-hand column are based on averaged data, and those in the middle and right-hand rows are based on individual subjects' data. The normalized spatial frequencies of s_1 and s_2 are indicated by gray bars. The panels in the upper two rows show data from the near conditions ($|s_1 - s_2| = 2$ just noticeable difference [JND] units); the panels in the lower two rows show data for the far conditions ($|s_1 - s_2| = 8$ JNDs).

captured by the model. Although the residuals are much larger in the individual-subject fits, one can see many of the same trends as are present in the model fits to the average data.

Separately fitting the four NEMO variants to each of the 10 subjects' data yielded a distribution of best-fitting parameter values across subjects. This analysis allowed us to generate between-subjects confidence intervals for each of the parameter values and, thereby, assess the variation in model parameters across subjects. Table 2 reports

the mean and standard error of the best-fitting parameter values for each of the four NEMO variants for the individual data.

By fitting all four model variants to individual subjects, we can quantitatively assess the differences in the BIC goodness-of-fit statistic for the candidate models. Table 3 shows the BIC values obtained for each subject and model. As can be seen, $NEMO_{\beta E}$ achieves the smallest BIC in 8 out of the 10 subjects. The advantage of $NEMO_{\beta E}$ over the other models proved significant ($p < .05$ by a permuta-

Table 2
Best-Fitting Parameter Values for NEMO's Fits to Results

Parameter	$NEMO_{\beta E}$		$NEMO_{\beta H}$		$NEMO_E$		$NEMO_H$	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
σ_1	2.109	0.166	2.346	0.200	2.299	0.239	2.364	0.218
σ_2	1.459	0.115	1.507	0.103	1.524	0.085	1.503	0.104
α	0.908	0.043	0.976	0.016	0.973	0.028	0.986	0.012
β	-0.755	0.117	-0.949	0.159				
τ	0.688	0.051			0.898	0.038		
Δ			1.897	0.099			1.773	0.075
Criterion	0.276	0.020	0.357	0.062	0.229	0.012	0.314	0.073
RMSD	0.113	0.005	0.121	0.005	0.129	0.005	0.125	0.005

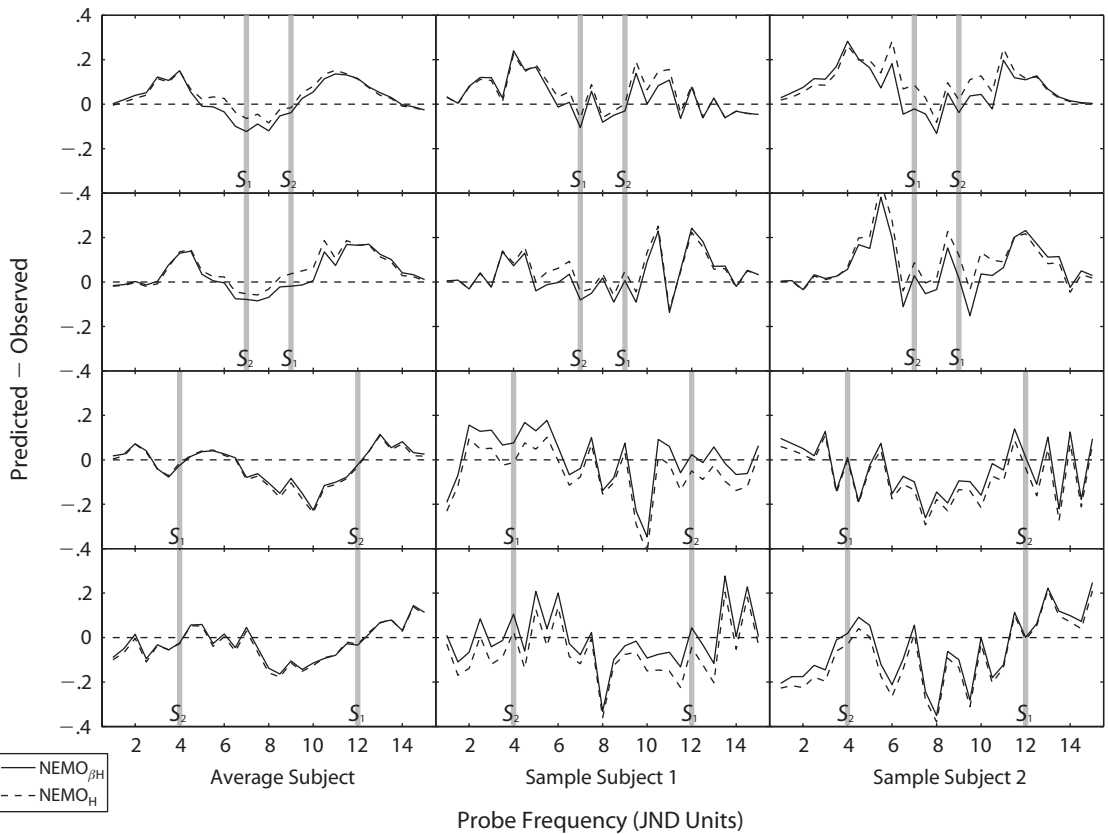


Figure 4. Difference between predicted and observed mnemometric functions for the Heaviside variants of NEMO with and without an adjustment for interstimulus similarity (i.e., $NEMO_{\beta H}$ and $NEMO_H$). Residuals in the left-hand column are based on averaged data, and those in the middle and right-hand rows are based on individual subjects' data. The normalized spatial frequencies of s_1 and s_2 are indicated by gray bars. The panels in the upper two rows show data from the near conditions ($|s_1 - s_2| = 2$ JNDs); the panels in the lower two rows show data for the far conditions ($|s_1 - s_2| = 8$ JNDs).

tion test, corrected for multiple comparisons). The other model comparisons did not attain statistical significance.

$NEMO_{\beta E}$'s fit indicated that interstimulus similarity moderated the effect of summed similarity on decisions. This phenomenon is seen in the significantly negative value of the β parameter (-0.76 ± 0.12), indicating that subjects are less likely to say *yes* when interstimulus similarity is high than when it is low. This extends previous findings of Kahana and Sekuler (2002) and Nosofsky and Kantner (2006) to an experimental paradigm in which only a single stimulus dimension (either horizontal or vertical frequency) varied.

NEMO has two mechanisms that enable it to fit the observed recency effects. First, NEMO can adopt different values for the standard deviation of ε for s_1 (denoted σ_1) and for s_2 (denoted σ_2). In fitting our data, NEMO consistently chose a smaller standard deviation for the memorial noise of the more recent stimulus (see Table 2). Borrowing from the literature on global-matching models of recognition memory (Murdock & Kahana, 1993b; Murdock & Lamon, 1988), we also allow NEMO to differentially weight the contribution of s_1 and s_2 in the summed-similarity calculation (this is done via the mod-

el's α parameter). This increased weighting of recently experienced items may be viewed either as a decay process or as a mechanism of strategic control. In the latter interpretation, subjects may focus their memory comparison on more recently presented information, downweighting the signals from older traces. In fitting our data, NEMO consistently chose to assign a greater weight to s_2 than to s_1 (see Table 2). One could easily imagine that the model

Table 3
Values of the Bayesian Information Criterion (BIC) for Each Subject and Model

Subject	$NEMO_{\beta E}$	$NEMO_{\beta H}$	$NEMO_E$	$NEMO_H$
1	-460.138	-408.299	-395.327	-398.468
2	-448.709	-429.838	-457.024	-435.387
3	-428.123	-419.197	-418.838	-424.400
4	-424.155	-395.422	-411.456	-403.301
5	-505.869	-485.721	-455.823	-473.885
6	-468.497	-453.988	-465.594	-458.012
7	-494.123	-480.834	-467.319	-484.200
8	-457.325	-450.624	-434.903	-447.940
9	-426.251	-422.282	-422.250	-427.311
10	-493.830	-442.971	-403.944	-420.087

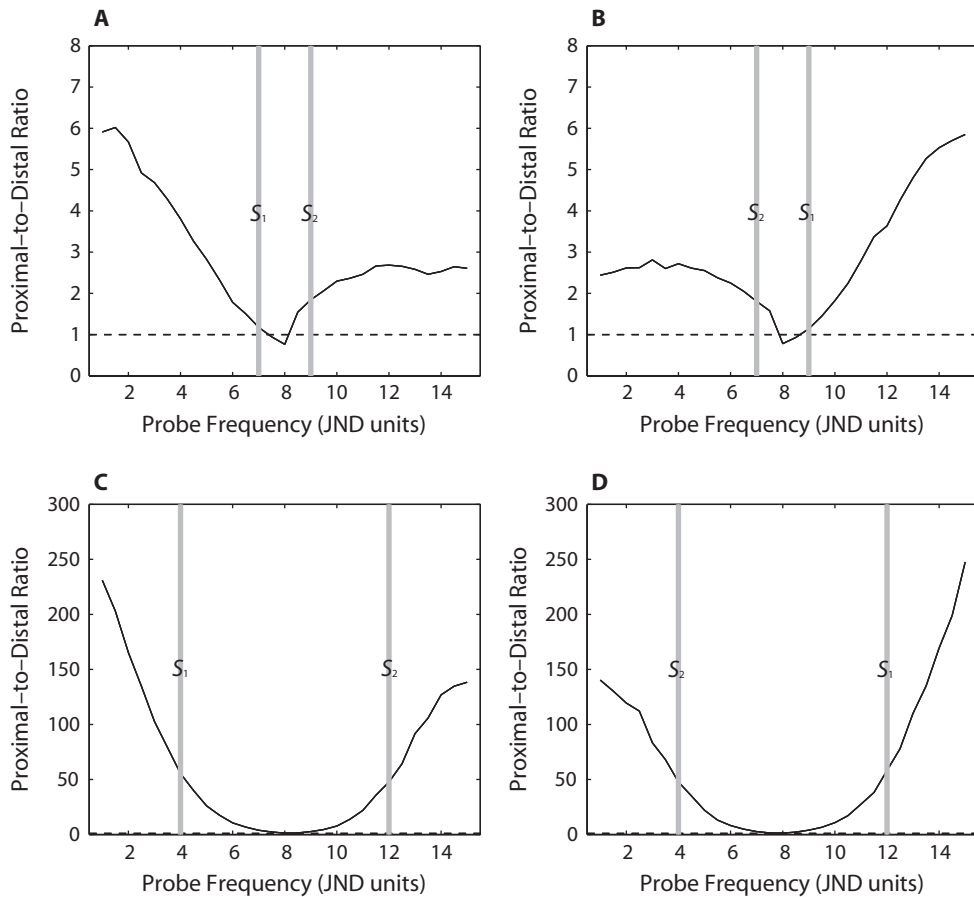


Figure 5. Relative contributions of proximal and distal items to summed similarity. Each panel shows p 's similarity to the proximal study item divided by p 's similarity to the distal study item as a function of probe position for one of the experimental conditions. The normalized spatial frequencies of s_1 and s_2 are indicated by gray bars. The upper panels show data from the near conditions; the lower panels show data from the far conditions. The panels on the left show data from the conditions in which s_1 had the lower spatial frequency of the two study items; the panels on the right show data from the conditions in which s_1 had the higher spatial frequency. The dashed line highlights the region of each curve where proximal and distal items contribute similarly to summed similarity. The values used are the mean over 5,000 repetitions of $\text{NEMO}_{\beta E}$, using the parameters given in Table 2.

would pick one of these two mechanisms over the other, at least for individual subjects. Yet both mechanisms appeared to play a role in accounting for the observed serial position effects. Given that the older stimulus is encoded with greater noise, it would be rational to assign it less weight in the summed-similarity computation.

Three factors conspire to determine how the proximal and distal stimuli (i.e., stimuli that were nearer to the probe on the spatial frequency dimension or further from the probe on that dimension) contribute to the overall summed-similarity calculation: (1) the form of the similarity–distance function as defined in Equation 2, (2) the noise added to the stored stimulus representations, and (3) the differential weighting of more and less recent items, as determined by α . When p 's spatial frequency is close to those of both s_1 and s_2 , the ratio of the summed-similarity contributions of proximal and distal items should approach one. However, when p 's spatial frequency is much nearer to one of the studied items, the

proximal item's contribution to summed similarity will be far greater than that of the distal item. Figure 5 shows how the ratio of the proximal and distal items' contributions to summed similarity changes as a function of p 's spatial frequency in each of the four experimental conditions. Here, one can see that the relative contribution of the proximal item to summed similarity, in comparison with that of the distal item, is substantially higher in the far condition (panels C and D) than in the near condition (panels A and B).

We next will consider what happens when we replace the exponential similarity–distance function used in $\text{NEMO}_{\beta E}$ with a Heaviside (step) function in $\text{NEMO}_{\beta H}$. Note first that even with the Heaviside function, NEMO chose a significantly negative value for the β parameter, indicating that when interstimulus similarity is low, subjects are more likely to say *yes* when the probe is similar to one of the studied items. Although Δ takes on a value smaller than the distance between the two studied items,

even in the near condition, the model still takes on a significantly negative value for β . This means that interstimulus similarity matters despite the fact that, on average, interstimulus similarity will be zero when the Heaviside function is used. This seems less mysterious when one considers that noise in the coding of s_1 and s_2 will occasionally make the distance between the studied items less than Δ ; therefore, interstimulus similarity will be 1 on a portion of the trials.

Our final manipulation was to eliminate the β parameter, yielding the models designated NEMO_E and NEMO_H. The effect of this manipulation can be seen in Figure 3 and Figure 4's plots of residuals (i.e., differences between predicted and observed mnemonic functions). In each panel, residuals associated with a model that includes β are shown as a solid line, and residuals associated with a model from which β has been eliminated are shown as a dotted line. In Figure 3, which presents results for the two exponential variants, NEMO _{β E} and NEMO_E, differences between residuals with and without β are more pronounced than the corresponding differences in Figure 4, which presents results for the model's two Heaviside variants, NEMO _{β H} and NEMO_H. These comparisons show clearly that the presence of β has a greater effect on the exponential models than on the Heaviside models.

Discussion

Summed-similarity models provide a major current framework for thinking about recognition and classification of stimuli whose similarity relations vary along measurable dimensions (e.g., Brockdorff & Lamberts, 2000; Estes, 1994; Nosofsky, 1991). An important contribution of these models has been in their ability to describe variation in false alarm rates as a function of the distance between a lure and the studied items (see also Hintzman, 1988).

Extending the summed-similarity analysis to account for data on individual lists, Kahana and Sekuler (2002) found that summed-similarity models mispredicted subjects' performance on lists with very high or very low interstimulus similarity. They also found that when lists were homogeneous, subjects committed fewer false alarms to similar lures. These findings, which were obtained using sinusoidal gratings as stimuli, were extended by Nosofsky and Kantner (2006) to Munsell color stimuli.

The present experiment examined the role of interstimulus similarity in predicting subjects' false alarms to lures varying systematically in their similarity to a list of two previously studied items. This parametric manipulation of the lures' perceptual coordinates reveals a functional relation between subjects' false alarm rates and the lure's similarity to the studied items. Zhou et al. (2004) referred to this relation as a mnemonic function and showed that a simple signal detection model could explain its basic form.

In the present study, we measured individual subjects' mnemonic functions for lists comprising two gratings that were separated by either 2 or 8 JNDs. Before testing memory, we measured individual subjects' discrimination thresholds and used those thresholds to scale or tailor stim-

uli for each subject. Armed with a substantial data set for recognition memory, we fit four variants of NEMO to individual subjects' mnemonic functions. The four models differed along two dimensions: (1) whether they used interstimulus similarity to adjust subjects' criteria on a trialwise basis and (2) whether they used an exponential or a Heaviside function to map perceptual distance (in JND units) onto psychological similarity. The study by Zhou et al. (2004) motivated us to evaluate the *form* of the similarity–distance function, because the version of NEMO that assumes a Heaviside function is quite similar to the signal detection framework presented in that article. Individual-subject fits demonstrated a clear superiority for Kahana and Sekuler's (2002) NEMO _{β E} model, which uses an exponential transformation of perceptual distance and uses interitem similarity to adjust the decision criterion (Nosofsky & Kantner, 2006). Note that despite NEMO _{β E}'s clear superiority, each of the model variants does a respectable job of fitting the mnemonic functions. This fact serves as a reminder that, as Pitt and Myung (2002) and others have cautioned, even a seemingly good fit can be bad.

We should note that NEMO includes some implicit assumptions that need to be tested explicitly. First, as presently formulated, the model treats each trial as encapsulated, with no intrusion or bleed-through from any preceding trial. Put another way, Equation 3 implies that the information used for each recognition decision is written on a blank slate, with memory being reset after each trial (cf. Howard & Kahana, 2002; Murdock & Kahana, 1993a). If this assumption were correct, recognition memory measured in this paradigm for grating stimuli would differ from memory for other classes of stimuli, including letters and words, which show strong interlist interference effects (e.g., Bennett, 1975; Donnelly, 1988; Monsell, 1978). Moreover, with a different paradigm, visual search experiments have demonstrated bleed-through from previous trials for features such as color (Huang, Holcombe, & Pashler, 2004) and, most directly relevant to our experiment, the spatial frequency of stimuli (Maljkovic & Nakayama, 1994). NEMO could be easily augmented to account for interlist interference effects, if needed to accommodate results from extensions of experiments such as the one we report here. Briefly, if recognition judgments on the n th trial were shown to be influenced by study items on the previous trial, NEMO could be augmented by an additional term that represents summed similarity between p on the n th trial and study items on the $(n-1)$ th trial. If the proportion of intertrial intrusions were modest, the term for intertrial summed similarity would be weighted substantially less than its intratrial counterpart.

Models of contextual drift (e.g., Dennis & Humphreys, 2001; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Murdock, 1997) offer a more sophisticated approach to the problem of interlist interference and contextually focused retrieval. Rather than downweighting the contribution of older memories to summed similarity, as we have done in NEMO, contextual drift models assume that each memory is tagged with a contextual representation that changes slowly over time. This vector representation of context, sometimes referred to as temporal

context, integrates the patterns of brain states over time, with more recent states weighted more heavily than older states (see Howard & Kahana, 2002, for details). Within this framework, a recognition test cue is a vector of attribute values whose elements include both the attributes representing the test item itself and the attributes representing the state of temporal context at the time of test (cf. Neath, 1993). This composite test cue, representing both content and context information, is used to retrieve similar traces stored in memory. The recognition decision is then based on the similarity of this retrieved information with the test cue itself (Dennis & Humphreys, 2001; Schwartz, Howard, Jing, & Kahana, 2005). Although context-based models have previously been applied primarily to data on verbal memory, it is possible to incorporate their contextual drift and contextual retrieval mechanisms into summed-perceptual-similarity models, such as NEMO.

The second assumption deserving of further study is the mechanism behind interitem similarity's influence on recognition. In NEMO, this influence is represented by the parameter β (see Equation 3). Fits of the model have produced consistently negative values for β , which is consistent with the assumption that intersimilarity leads to an adaptive shift in the subject's criterion (Kahana & Sekuler, 2002; Nosofsky & Kantner, 2006). When study items are highly similar to one another, the associated negative β seems to increase the effective conservatism of the subject's criterion for responding *yes*. Hence, the negative β affords some protection against false positives. This interpretation, which locates β 's role at a decision-making stage, seems a reasonable hypothesis, but one that future research might examine. Several physiological and behavioral studies hint at an alternative worth considering—namely, that β is mediated by a change in the sensitivity or noise of early visual mechanisms in cortical processing, well before the decision-making stage. For example, Spitzer, Desimone, and Moran (1988) showed that the difficulty level of a visual discrimination task altered neuronal responses in visual area V4 of the cerebral cortex. Presumably, a more difficult task makes for increased attentional demands, which increase the strength and selectivity of neuronal responses (see also Boynton, 2005; McAdams & Maunsell, 1999; Reynolds & Desimone, 2003). It would be useful, therefore, to determine whether, in recognition memory, interitem similarity has an analogous effect, altering the noise or sensitivity of visual mechanisms responsible for the encoding of study items.

AUTHOR NOTE

The empirical study on which this report is based was designed and carried out by F.Z. as part of his doctoral research at Brandeis. The authors acknowledge support from National Institutes of Health Grants MH55687, MH062196, and MH68404, from NSF Grant 0354378, and from U.S. Air Force Office of Scientific Research Grant F49620-03-1-037. We also thank Marieke van Vugt for her helpful comments on a previous draft. Correspondence concerning this article may be addressed to M. J. Kahana, Department of Psychology, University of Pennsylvania, 3401C Walnut St., Room 302C, Philadelphia, PA 19104 (e-mail: kahana@psych.upenn.edu) or to F. Zhou or R. Sekuler, Center for Complex Systems, Brandeis University, Waltham, MA 02454 (e-mail: fzhou@brandeis.edu or sekuler@brandeis.edu).

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* (pp. 251-301). Mahwah, NJ: Erlbaum.
- BENNETT, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning & Verbal Behavior*, *14*, 573-584.
- BOYNTON, G. M. (2005). Attention and visual perception. *Current Opinion in Neurobiology*, *15*, 465-469.
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.
- BROCKDORFF, N., & LAMBERTS, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 77-102.
- CHATER, N., & VITÁNYI, P. M. B. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, *47*, 346-369.
- 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.
- DELLA-MAGGIORE, V., SEKULER, A. B., GRADY, C. L., BENNETT, P. J., SEKULER, R., & MCINTOSH, A. R. (2000). Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *Journal of Neuroscience*, *20*, 8410-8416.
- DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- DONNELLY, R. E. (1988). Priming effects in successive episodic tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 256-265.
- ENNIS, D. M., PALEN, J., & MULLEN, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*, 449-465.
- ESTES, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in multiple-trace memory model. *Psychological Review*, *95*, 528-551.
- HOWARD, M. W., & KAHANA, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269-299.
- HUANG, L., HOLCOMBE, A. O., & PASHLER, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, *32*, 12-20.
- HUMPHREYS, M. S., PIKE, R., BAIN, J. D., & TEHAN, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*, 36-67.
- HWANG, G., JACOBS, J., GELLER, A., DANKER, J., SEKULER, R., & KAHANA, M. J. (2005). EEG correlates of verbal and nonverbal working memory. *Behavioral & Brain Functions*, *1*, 20.
- KAHANA, M. J., & BENNETT, P. J. (1994). Classification and perceived similarity of compound gratings that differ in relative spatial phase. *Perception & Psychophysics*, *55*, 642-656.
- KAHANA, M. J., RIZZUTO, D. S., & SCHNEIDER, A. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 933-953.
- KAHANA, M. J., & SEKULER, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*, 2177-2192.
- LAMBERTS, K., BROCKDORFF, N., & HEIT, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General*, *132*, 351-378.
- MADDOX, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics*, *61*, 354-374.
- MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 795-817.
- MALJKOVIC, V., & NAKAYAMA, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*, 657-672.
- MCADAMS, C. J., & MAUNSELL, J. H. (1999). Effects of attention on

- the reliability of individual neurons in monkey visual cortex. *Neuron*, **23**, 765-773.
- McKINLEY, S. C., & NOSOFSKY, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 294-317.
- MENSINK, G. J., & RAAIJMAKERS, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, **95**, 434-455.
- MONSELL, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, **10**, 465-501.
- MORGAN, M. J., WATAMANIUK, S. N. J., & MCKEE, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, **40**, 2341-2349.
- MURDOCK, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, **104**, 839-862.
- MURDOCK, B. B., & KAHANA, M. J. (1993a). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 689-697.
- MURDOCK, B. B., & KAHANA, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1450-1453.
- MURDOCK, B. B., & LAMON, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition*, **16**, 91-101.
- NEATH, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, **21**, 689-698.
- NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multi-dimensional models of perception and cognition* (pp. 363-394). Hillsdale, NJ: Erlbaum.
- NOSOFSKY, R. M., & KANTNER, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition*, **34**, 112-124.
- 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). Designing a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, **2**, 187-199.
- PETZOLD, P., & HAUBENSAK, G. (2004). Short-term and long-term frames of reference in category judgments: A multiple-standards model. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 45-68). Mahwah, NJ: Erlbaum.
- PITT, M. A., & MYUNG, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, **6**, 421-425.
- REYNOLDS, J. H., & DESIMONE, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, **37**, 853-863.
- RIZZUTO, D. S., & KAHANA, M. J. (2001). An autoassociative neural network model of paired-associate learning. *Neural Computation*, **13**, 2075-2092.
- SCHWARTZ, G., HOWARD, M. W., JING, B., & KAHANA, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, **16**, 898-904.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- SEKULER, R., KAHANA, M. J., McLAUGHLIN, C., GOLOMB, J., & WINGFIELD, A. (2005). Preservation of episodic visual recognition memory in aging. *Experimental Aging Research*, **31**, 1-13.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SPITZER, H., DESIMONE, R., & MORAN, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, **240**, 338-340.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.
- STERNBERG, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **27**, 1-32.
- TREISMAN, M., & WILLIAMS, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, **91**, 68-111.
- WETHERILL, G. B., & LEVITT, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical & Statistical Psychology*, **18**, 1-10.
- 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.

NOTES

1. Zhou et al.'s (2004) signal detection model assumed that *yes* decisions arose when either the absolute difference between the remembered exemplar s_1 and the probe or the absolute difference between the remembered exemplar s_2 and the probe was less than a threshold value. Thus, Zhou et al.'s model would substantively differ from NEMO_H only in those cases in which both study items were sufficiently proximate to the probe item to produce a nonzero similarity value.

2. We used an evolutionary algorithm to search for the parameter values that minimized the RMSD (for further details on parameter estimation, see Kahana, Rizzuto, & Schneider, 2005; Kahana & Sekuler, 2002; Rizzuto & Kahana, 2001).

(Manuscript received October 23, 2005;
revision accepted for publication July 17, 2006.)