An Electrophysiological Signature of Summed Similarity in Visual Working Memory

Marieke K. van Vugt
University of Groningen

Robert Sekuler
Brandeis University

Hugh R. Wilson
York University

Michael J. Kahana
University of Pennsylvania

Summed-similarity models of short-term item recognition posit that participants base their judgments of an item's prior occurrence on that item's summed similarity to the ensemble of items on the remembered list. We examined the neural predictions of these models in 3 short-term recognition memory experiments using electrocorticographic/depth electrode recordings and scalp electroencephalography. On each experimental trial, participants judged whether a test face had been among a small set of recently studied faces. Consistent with summed-similarity theory, participants' tendency to endorse a test item increased as a function of its summed similarity to the items on the just-studied list. To characterize this behavioral effect of summed similarity, we successfully fit a summed-similarity model to individual participant data from each experiment. Using the parameters determined from fitting the summed-similarity model to the behavioral data, we examined the relation between summed similarity and brain activity. We found that 4–9 Hz theta activity in the medial temporal lobe and 2–4 Hz delta activity recorded from frontal and parietal cortices increased with summed similarity. These findings demonstrate direct neural correlates of the similarity computations that form the foundation of several major cognitive theories of human recognition memory.

Keywords: summed similarity, oscillations, recognition memory, EEG

Behavioral and model-theoretic analyses of recognition memory agree that similarity is a major determinant of both recognition accuracy and speed (e.g., Hintzman, 1988; Nosofsky, Little, Donkin, & Fific, 2011). Specifically, participants' recognition responses are influenced by a probe's similarity to items on the just-studied list. When a lure is similar to one or more studied items, participants tend to commit false alarms, and when they correctly reject those similar items, their response times (RTs) are significantly increased. Moreover, a target is more likely to be endorsed when it is similar to the nontarget items on the studied list (see Kahana, 2012, for a review).

Assessing the psychological similarity of a probe (be it a target or lure) to the stored representation of the items on a just-presented list requires a quantitative model of the mapping between physical and psychological coordinates and of the way in which similarities combine across multiple study items. Summed-similarity models, which describe both of these processes, have been highly successful in accounting for behavioral data on item recognition (e.g., Hintzman, 1988; Kahana & Sekuler, 2002; Lamberts, Brockdorff, & Heit, 2003; Nosofsky, 1991; Nosofsky & Kantner, 2006; Nosofsky et al., 2011; Shiffrin & Steyvers, 1997). These models posit that item recognition is a signal-detection process, in which a participant computes the probe's summed similarity to all study items and bases her or his recognition decision on this measure. Whenever the computed summed similarity exceeds a decision threshold, the participant will respond *yes.* Thus, when summed similarity is high for probes that are targets, there is little interference and the participant is likely to respond *yes.* For probes that are lures, high summed similarity constitutes high interference, which also increases the participant's tendency to incorrectly respond *yes.*

Summed-similarity models are generally viewed as models of familiarity. There is evidence that people can base recognition judgments both on recall of specific features of the encoding trial (termed "recollection") and on item familiarity. However, for lists of perceptually similar items, pure summed-similarity (familiarity) models provide an excellent fit to the data without positing a recollection component. Given summed-similarity models' success in describing trial-by-trial variation in participants' behavior

(e.g., the noisy exemplar model; NEMo; Kahana & Sekuler, 2002; Nosofsky & Kantner, 2006), we sought to determine whether there is an electrophysiological correlate of the model-defined summed-similarity signal. Such a demonstration would provide neural support for the existence of the cognitive construct of summed similarity.

In three studies we examined whether oscillatory activity in both electrocorticographic (ECoG) recordings in neurosurgical patients and scalp EEG recordings in healthy controls correlate with summed-similarity computations, with other relevant variables being controlled for. We focused our attention on oscillatory responses because an extensive body of work has linked neural oscillations to cognitive processes in both humans and animals (see Jacobs & Kahana, 2010, for a review). Of particular relevance to the current studies is the finding that during memory retrieval, slow oscillations, in the theta (4–9 Hz) and delta (2–4 Hz) frequency bands, are of higher amplitude for targets than for lures (Düzel, Neufang, & Heinze, 2005; Jacobs, Hwang, Curran, & Kahana, 2006) and also of higher amplitude for high-confidence correct decisions than for low-confidence correct decisions (Guderian & Düzel, 2005). In contrast, oscillations at higher frequencies, specifically in the gamma band (28–128 Hz), have primarily been associated with manipulations of memory load (Howard et al., 2003; van Vugt, Schulze-Bonhage, Litt, Brandt, & Kahana, 2010). The ECoG study reported here builds on a prior study by van Vugt et al. (2009), which involved a subset of the same group of neurosurgical patients. By combining novel analyses of those ECoG data with data from two new scalp EEG studies we assayed neural data recorded at multiple spatial scales for direct evidence of summed-similarity computations. To parametrically investigate the neural correlates of similarity in item recognition, we used specially designed synthetic faces whose similarity is well defined both in physical and psychological representational spaces (see Method below).

Experiments 1–3

Method

Because similar experimental and data-analytic methods were used in Experiments 1–3, we present the methods for all three experiments together and note differences between studies as appropriate.

Synthetic faces. We chose human faces as stimuli because of their ecological validity and because their similarities can be both quantified and parametrically varied. Rather than using photographs of faces as stimuli, we adopted a set of synthetic faces derived from real photographs but filtered to remove components such as hair and skin texture while preserving the shape, size, and position of key facial features (Wilson, Loffler, & Wilkinson, 2002). Wilson's technique for generating synthetic faces begins by taking 37 physical measurements on a set of (normalized) photographs of Caucasian males and then reconstructing the faces from the principal components that could be extracted from the matrix of these measurements. A stimulus set of 16 faces was constructed from all permutations of steps of one standard deviation away from the mean face (of the 37-dimensional face space) in the directions of each of the first four principal components (one standard deviation is approximately the threshold for 75% correct discrimina-

tion between two faces that are flashed for 110 ms; see Appendix A). The resulting faces were low-pass filtered at 10 cycles per face width, which is optimal for face processing and removes components like hair and skin texture (Wilson et al., 2002). The advantages of using this face set are that, although the faces are well controlled, they still can be identified with high accuracy, and it is possible to measure their interitem similarity precisely. Moreover, the faces are realistic enough to generate strong responses in the fusiform face area (Loffler, Yourganov, Wilkinson, & Wilson, 2005). This is important because previous studies have shown that the fusiform face area is sensitive to similarity distances between faces (e.g., Harris & Aguirre, 2010). As such, it should be possible to find neural effects of similarity for these faces.

Multidimensional scaling study. We carried out a preliminary multidimensional scaling (MDS) study to characterize the psychological similarities among the items (see Appendix A for details). In this MDS study, 23 participants served for two sessions, in which they saw all combinations of three faces (triads) twice and had to determine the "odd-one-out" (i.e., the face that is least similar to the other two). From these ratings, a similarity matrix was constructed by increasing the similarity value for each of the two nonchosen faces (see Kahana & Bennett, 1994, for details). This similarity matrix was transformed into similarity coordinates for every face using individual-differences multidimensional scaling (INDSCAL/ALSCAL; Takane, Young, & de Leeuw, 1977). We used the four-dimensional stimulus coordinates derived from this procedure (see Appendix A) in our summed-similarity analyses. The normalized raw stress measure for this solution was 0.022, which had a dispersion accounted for (DAF) of 0.98. (DAF is analogous to variance accounted for in a regression analysis; Heiser & Bennani, 1997.) The attentional weights were very similar across participants, as demonstrated by the small standard deviation for each of the four dimensions: mean attentional weight (standard deviation) = 0.34 (0.026); 0.34 (0.042); 0.34 (0.024); 0.33 (0.015).

Short-term item recognition paradigm. Figure 1 illustrates the sequence of events during each trial of the experiment. Following the appearance of a fixation stimulus (an asterisk that appeared for 1,000–1075 ms, jittered) participants viewed a short series of faces. Each stimulus appeared for 700–775 ms, followed by a 275- to 350-ms interstimulus interval. After a 3,000- to 3,300-ms retention interval, a probe item appeared and participants indicated with a keypress whether the probe was a member of the just-presented list (a target) or an item not shown on the current list (a lure). After each trial, participants were given accuracy feedback. The participant advanced to the next trial with a keypress. Temporal jitter was used to avoid spurious correlations between ongoing brain activity and task events. Participants were encouraged to be both fast and accurate, and participants in Experiments 2 and 3 were paid a bonus based on their accuracy and response time. The experiment was programmed in the Python Experimental Programming Library PyEPL (http://pyepl.sourceforge.net; described in Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

The three experiments differed in the number of faces presented in each series. Lists used in Experiments 1 and 2 comprised a variable number of faces, whereas lists in Experiment 3 always comprised three faces. In Experiment 1, lists could have one, two, or three faces; in Experiment 2, lists could have two, three, or four
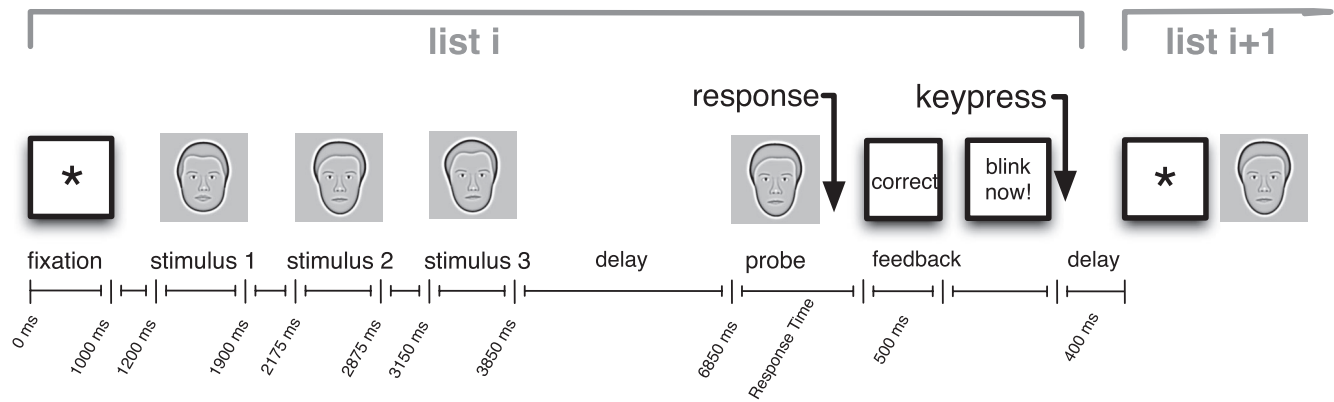
Figure 1. Trial structure of the Sternberg task. This figure illustrates the sequence and timing of events in a trial with three faces.

faces. Experiments 1 and 2 also contained lists of nonface stimuli, which are not analyzed in the present report.[1]

Each experiment employed a "recent negatives" manipulation to study the effects of proactive interference on brain activity (Monsell, 1978). Whenever a lure probe was shown, it always had been a study item one, two, or three lists back (we refer to this lure recency as "list lag"). Lures that appeared as study items on recent lists are expected to give rise to significant proactive interference. Although we report on the behavioral effects of lure recency, these effects are not a major focus of the present study. With the exception of those probe items that were part of the recent negatives manipulation, lists were constructed so that items could not be repeated on successive lists. Lists were also constructed so that the frequencies of targets and lures and each of the list lengths and the three levels of the recent negatives manipulation were all matched. In addition, targets were equally likely to match a study item from each serial position. Blocks were then constructed by drawing lists at random from the pool, subject to the above constraints.

Every session was preceded by two 16-trial training blocks; for Experiment 1, there were 40 additional one-item lists to familiarize the participant with the face stimuli. Participants were given feedback on their average accuracy and RT at the end of each 30-trial block. Incorrect trials and trials with RTs shorter than 200 ms or longer than 2,500 ms (for Experiment 1) or 2,000 ms (for Experiments 2 and 3) were removed from the analysis.

Participants. Participants in Experiment 1 were 16 neurosurgical patients being treated for medically refractory epilepsy (ages 15–58; six female). Patients were monitored with arrays of subdural and/or depth electrodes to localize seizure onset and to map cognitive functions prior to resective surgery. Patients were recruited from Brigham and Women's Hospital (Boston, Massachusetts), the Hospital of the University of Pennsylvania (Philadelphia), and Universitäts Klinikum Freiburg (Freiburg, Germany), and the research protocol was approved by the appropriate institutional review boards at each hospital. Informed consent was obtained from all participants.

In Experiment 2, 37 young adults (ages 20–32) recruited from the University of Pennsylvania student community participated in two experimental sessions for payment. Each session involved 10 blocks of 30 trials, as described above. In Experiment 3, 20 young

adults (ages 19–29) participated in four experimental sessions for payment. Over the course of their four sessions, participants saw the same set of 66 study lists 16 times. The lists were divided into three blocks, the order of which was randomized across the four repetitions in each testing session.

ECoG and depth electrode recordings in Experiment 1. The local field potential was amplified and digitally recorded at sampling rates between 250 and 1,024 Hz and bandpass filtered between 0.1 and 100 Hz. Data were subsequently notch filtered with a Butterworth filter with zero phase distortion to eliminate line noise. The intervals of interest were scanned for epileptic spikes and sharp waves by means of a kurtosis threshold; events were discarded if their kurtosis exceeded a threshold of 5 (see van Vugt et al., 2010, for details).

To synchronize electrophysiological recordings with behavioral events, the computer generating the task sent pulses via an optical isolator into an unused recording channel or digital input on the amplifier to time stamp the digital intracranial EEG (iEEG) recording. The time stamps associated with these pulses aligned the experimental computer's clock with the iEEG clock to a precision well under the sampling interval of the iEEG recording (<4 ms). For all participants, the locations of implanted electrodes were determined by means of co-registered postoperative computed tomographies and preoperative magnetic resonance imaging (MRI) or from postoperative MRIs by an indirect stereotactic technique and converted into Montreal Neurological Institute (MNI) coordinates. Localization of depth electrode contacts in the hippocampus was done manually through clinician's inspection of the postoperative MRIs.

Scalp EEG recordings in Experiments 2 and 3. We recorded scalp EEG signals using a 129-channel EGI Inc. system, with an AC-coupled, high-input-impedance amplifier (200 MΩ, Net Amps, Electrical Geodesics, Inc., Eugene, OR). The sampling rate was 500 Hz, and data were recorded with a 0.1–250 Hz bandpass filter. Individual channels were adjusted until imped-

---

[1] Experiment 1 included lists of consonants, and Experiment 2 included lists of compound sinusoidal gratings. For these experiments, trials of each list type were presented in blocks of 30 lists with the order of the blocks randomized across participants.

ances were below 50 kΩ. Bad channels were detected manually off-line, and the remaining channels were used for average re-referencing. Eye blinks were detected using a running-average threshold and a voltage threshold of 100 μV. Data from seven participants for whom more than 20% of events were rejected due to eye blinks were excluded, leaving us with 30 participants. One additional participant was excluded because of average accuracy below 60%. For Experiment 3, one participant was excluded because of average accuracy below 60%, and another was excluded for failure to exhibit serial position effects (and accuracy below 70%). All data were notch filtered in the 58–62 Hz frequency band to remove line noise.

Regression-based analysis of EEG and ECoG oscillations. We used a multiple-regression approach (e.g., Jacobs et al., 2006) to assess the neural correlates of summed similarity while controlling for the effects of proactive interference and within-list recency (i.e., serial position effects that cause the first and last items of a list to be remembered better). We computed the ordinary least squares regression coefficient of summed-similarity S, response time, and recency on z-transformed oscillatory power during the interval between probe onset and the moment the participant responded. Oscillatory power was determined using six-cycle Morlet wavelets at logarithmically spaced frequencies ($2^{(x/8)}$ Hz for $x \in 8, \ldots, 56$). Recency was defined separately for targets (lag_T; in units of item position) and lures (lag_L; the number of lists ago the probe was last presented). Note that lure recency is a measure of proactive interference, with more recent lures corresponding to conditions of higher proactive interference. S was computed by fitting a summed-similarity model (Kahana & Sekuler, 2002, described below) to the behavioral data from each individual participant. The parameters of this model were optimized with a pattern search algorithm that searched for the minimum of a goodness-of-fit measure that consisted of the sum of the root-mean-square deviation between modeled and empirical summed similarity and the difference between modeled and empirical d prime. In Experiments 1 and 2, S was binned into five levels of summed similarity; in Experiment 3, we had multiple repetitions of each list so we could fit the model to participants' accuracy data on each individual list.

To assess whether the neural correlates of S changed over the course of the trial and/or differed between low- and high-frequency bands, we included terms representing the interaction between S and frequency band F (2–4 Hz delta, 4–9 Hz theta, and 9–14 Hz alpha were coded as low frequencies; 14–28 Hz beta and 28–90 Hz gamma were coded as high frequencies) and between S and time interval T (first half or second half of the recognition probe-to-response interval). We also included the triple interaction term between S, F, and T. The length of the probe interval varied from trial to trial as a function of the participant's RT (i.e., oscillatory power was Vincentized; it was resampled into time bins whose size scaled with the length of the trial such that each trial had two time bins, but their size increased with the participant's response time; Workman & Adams, 1950). Because recency was defined separately for targets and lures, we fit separate models to the electrophysiological data recorded during these two types of trials. For target trials we fit the model

$$Z(P)_e = \beta_0 + \beta_1 S + \beta_2 lag_T + \beta_3 S \times F + \beta_4 S \times T$$
$$+ \beta_5 S \times F \times T + \beta_6 RT,$$

and for lure trials we fit the model

$$Z(P)_e = \beta_0 + \beta_1 S + \beta_2 lag_L + \beta_3 S \times F + \beta_4 S \times T$$
$$+ \beta_5 S \times F \times T + \beta_6 RT,$$

where $Z(P)_e$ indicates the z-transformed power at electrode e.

The resulting regression coefficients were then submitted to an across-subject t test to establish population effects. For Experiments 2 and 3, we used a cluster-based method on the t statistics to correct for multiple comparisons (Maris & Oostenveld, 2007). Clusters reflect the intuition that results are more likely to be true if they are present in groups of adjacent EEG channels. The procedure starts by comparing the p value from each channel's t test (on the regression coefficients) to a threshold of 0.01. We then assessed to what extent this binarized significance map consisted of clusters of spatially adjacent channels. For each cluster, we summed the t statistics from all the channels comprising that cluster, which we compared to a permutation distribution to determine whether the cluster was significant. The permutation distribution was created by randomly swapping the sign of individual-subject regression coefficients and recomputing the clusters. Significant clusters are those whose summed t statistic is in the upper tail of the distribution of summed t statistics based on randomized data. For Experiment 1, no such clustering was necessary, because we restricted our analyses to a small set of pre-defined regions of interest (ROIs).

The ROIs (predefined for Experiment 1; based on the $S \times F$ interaction for Experiments 2 and 3) were then used to examine the slope of oscillatory power on S. We used this $S \times F$ interaction to define our ROIs because we were primarily interested in oscillatory correlates of S that were specific for a frequency band. To ensure that any differences we would observe between Experiments 2 and 3 were not an artifact of having ROIs comprising different sets of electrodes, we defined a parietal and frontal ROI based on the intersection of the ROIs for the two experiments in the lures condition. For each of these ROIs, we then assessed whether the slope of S on oscillatory power was consistent between target and lure trials and between the two experiments.

Results

Behavioral analyses

Serial position effects. We observed standard serial position effects in all three experiments, with the general pattern being a large recency effect and a small primacy effect. Serial position had a significant effect on accuracy but not RT in Experiment 1, as revealed by a repeated-measures analysis of variance: accuracy, $F(2, 44) = 5.67$, $p < .01$; RT, $F(2, 44) = 1.29$, ns. In Experiment 2 we observed significant serial position effects for both accuracy and RT: accuracy, $F(3, 84) = 23.5$, $p < .001$; RT, $F(3, 84) = 8.0$, $p < .001$. In Experiment 3 we also observed reliable serial position effects for both accuracy and RT: accuracy, $F(2, 53) = 16.9$, $p < .001$; RT, $F(2, 53) = 35.6$, $p < .001$. For these analyses of variance we considered the longest list length in each of the experiments, which was list length 4 in Experiment 2 and list length 3 in

Experiments 1 and 3. Similar results were observed for list length 3 in Experiment 2.

Proactive interference effects. We next examined whether recently studied lures (i.e., lures that were members of recent prior lists) were more likely to be endorsed as targets. Although we did not observe a significant effect of our proactive interference manipulation in Experiment 1, we did observe clear proactive interference effects in both Experiment 2 and Experiment 3. In Experiment 2, both RT and accuracy showed a significant proactive interference effect. We quantified this proactive interference effect separately for each participant by regressing his or her accuracy on list lag (where a list lag of 1 corresponds to high proactive interference and a list lag of 3 corresponds to low proactive interference). The mean slope of this regression (0.028) was significantly larger than zero, $t(28) = 3.31$, $p < .001$. The mean slope of RT (-27.7 ms) exhibited an analogous effect, $t(28) = -3.72$, $p < .001$. In Experiment 3, we found a significant proactive interference effect for accuracy; the mean slope of accuracy on the list lag of lures was 0.063, $t(17) = 8.5$, $p < .001$, but there was no effect of lure lag on RT (mean slope $= 0.59$ ms), $t(17) = 0.14$, *ns*.

Application of summed-similarity model. We hypothesized that the basic computation underlying short-term item recognition is the summed pairwise similarity between an item's memorial representation and the representation of the test probe (Kahana & Sekuler, 2002; Nosofsky et al., 2011). 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. Kahana and Sekuler (2002) extended the basic summed-similarity framework for item recognition (Hintzman, 1988; Nosofsky, 1991) by showing that the degree of summed similarity required to respond *yes* varies with list homogeneity (i.e., the mean similarity among list items). With greater list homogeneity, participants will respond *yes* at lower levels of summed similarity (Kahana, Zhou, Geller, & Sekuler, 2007; Nosofsky & Kantner, 2006; Visscher, Kaplan, Kahana, & Sekuler, 2007). Specifically, according to Kahana and Sekuler's noisy exemplar model (NEMo), participants respond *yes* when the similarity signal (S) defined as

$$S = \sum_{i=1}^{LL} \alpha_i e^{-\tau|\mathbf{s_i}-\mathbf{p}|} + \frac{2\beta}{LL(LL-1)} \sum_{i=1}^{LL-1} \sum_{j=i+1}^{LL} e^{-\tau|\mathbf{s_i}-\mathbf{s_j}|} \quad (1)$$

exceeds a decision threshold c_{LL} (the subscript LL indicates there is a separate optimal decision criterion for every list length). In Equation 1, LL is list length, α_i is a forgetting parameter, τ determines the form of the exponential generalization gradient (Shepard, 1987), $\mathbf{s_i}$ is the vector representing the coordinates of the memorial representation of stimulus i as determined by a prior MDS study, $\mathbf{p}$ is the representation of the probe item, and β determines the influence of list homogeneity (bold typeface indicates vectors). A Gaussian random variable with mean zero and standard deviation σ is added to each item's memorial representation to simulate the variability in participants' responses across different occurrences of the same item.

Our implementation is a reduced form version of previous versions of NEMo insofar as we assume that σ does not vary across stimulus dimensions and that each dimension is weighted equally in the similarity computation. We also assume that the decision threshold is chosen to yield the maximum likelihood of

the simulated data. Although NEMo does not offer a process account of primacy or recency effects, the set of parameters α_i ($i = 1$ indicating the most recently studied item) allows the model to differentially weight items from different serial positions, thereby matching any forgetting or primacy effects that may be present in the data (see also Nosofsky et al., 2011).

Although most previous applications of summed-similarity theory have involved fitting group data, we chose to fit the model separately to each individual's behavioral data. In this way our aim of examining the neural correlates of summed similarity would not be subject to model misspecification at the level of individual participants. This approach enabled us to examine the relation between brain activity and summed similarity at the individual subject level and then to aggregate those tests across participants.

Estimates of the best fitting parameters we obtained (see Table 1) are similar across the three experiments and also similar to those reported in previous applications of NEMo (Kahana & Sekuler, 2002; Kahana et al., 2007; Nosofsky & Kantner, 2006; Visscher et al., 2007; Yotsumoto, Kahana, Wilson, & Sekuler, 2007). Not surprisingly, the alpha parameters mirror the serial position effects in the data, with a large recency effect and a small primacy effect across studies. The list homogeneity parameter, β, was negative in all studies and significantly negative in Experiments 2 and 3, which had the most statistical power, $t(28) = -2.295$, $p < .05$ and $t(17) = -5.58$, $p < .01$ for Experiments 2 and 3, respectively. This indicates that in general, list homogeneity decreases the decision threshold for responding *yes*.

The central tenet of summed-similarity theory is that the probability of responding *yes* to lures and to targets should increase with summed similarity. Figure 2 shows this effect in each of the three experiments. In all of the experiments, the false alarm rate reliably increased with summed similarity: Experiment 1: $M = 4.75$, $t(15) = 4.06$, $p < .01$; Experiment 2: $M = 6.15$, $t(28) = 7.94$, $p < .001$; Experiment 3: $M = 6.56$, $t(17) = 6.23$, $p < .001$. For targets, the effect was generally smaller, but nonetheless the hit rate was statistically significant in both Experiment 1, $M = 2.32$, $t(15) = 3.22$, $p < .01$, and Experiment 3, $M = 1.66$, $t(17) = 4.72$, $p < .001$.

Electrophysiological results. To avoid contaminating any observed interference-based effects with effects caused by differences in accuracy, we restricted our analyses to correct trials. For lure trials, where proactive interference was explicitly manipu-

Table 1
Best Fitting Parameter Values for NEMo's Fit to Behavioral Data in Experiments 1–3

Parameter	Experiment 1		Experiment 2		Experiment 3	
	M	*SEM*	*M*	*SEM*	*M*	*SEM*
β	-0.31	0.52	-1.06	0.46	-1.68	0.30
τ	1.32	0.17	1.65	0.12	1.89	0.16
σ	1.21	0.084	0.99	0.061	0.82	0.048
α_2	0.58	0.041	0.76	0.038	0.71	0.032
α_3	0.67	0.051	0.65	0.042	0.81	0.033
α_4			0.72	0.038		
RMSD	0.047	0.0026	0.021	0.0022	0.0811	0.0040

Note. NEMo = noisy exemplar model; RMSD = root-mean-square deviation; *SEM* = standard error of the mean.
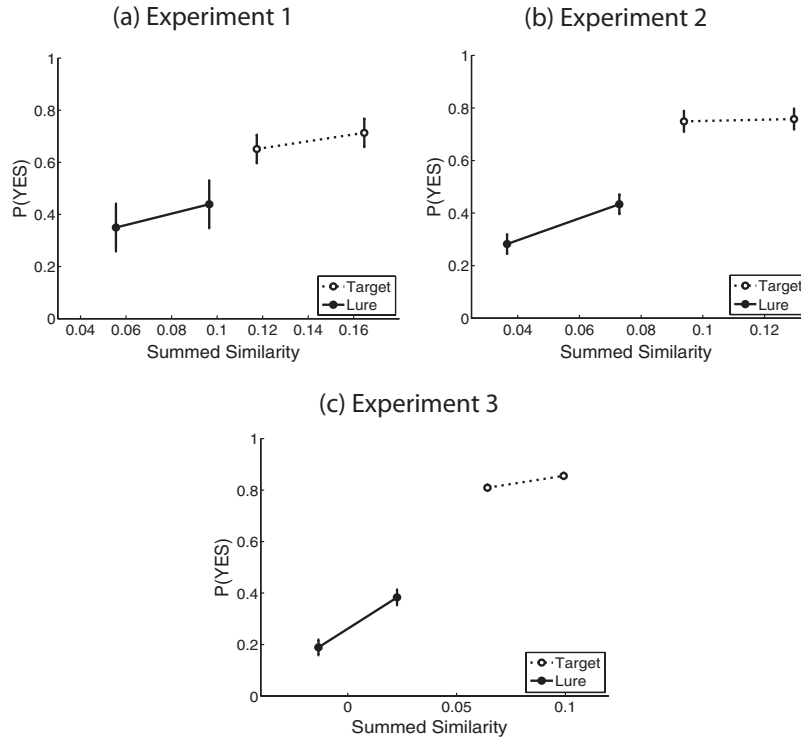
Figure 2. Summed-similarity effect. The probability (P) of saying *yes* is shown as a function of binned summed similarity (defined in Equation 1) for both targets (open circles) and lures (filled circles). Bins are defined as the lowest and highest 50% of the summed-similarity distribution. Error bars reflect Loftus–Masson corrected 95% confidence intervals. The fits for Experiment 1 are based on summed similarity for list length 3 only to make it comparable to the fits for Experiment 2, in which there were only list length 3 trials.

lated, we regressed oscillatory power at each electrode on summed similarity, lure recency (a measure of proactive interference), and RT. For target trials, where proactive interference was not manipulated, we regressed oscillatory power on summed similarity, target item recency, and RT. We fit the regression models (see Method) to each individual's data, using the NEMo fits to determine the value of S on each trial. We then asked whether the distribution of regression coefficients across participants deviated significantly from zero and whether this difference was modulated by time interval (first half vs. last half of retrieval period), frequency range (low vs. high), or both. On the basis of previous studies of the target-lure effect (e.g., Düzel et al., 2005; Jacobs et al., 2006), we expected to find a neural correlate of summed similarity in the lower frequency bands (particularly delta and theta).

For our ECoG and depth electrode recordings in Experiment 1, we computed an average regression coefficient for every participant across all the electrodes in a set of five predefined regions (see Table 2). Statistical significance was determined using a *p*-value threshold set by a false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). A FDR of 0.05, as used here, means that one expects to find 5% of the effects deemed significant to be false positives. In our study, a FDR of 0.05 corresponded to a conventional *p*-value threshold of .011.

We would expect an oscillatory correlate of summed similarity to be present for both targets and lures. Whereas a main effect of summed similarity, *S*, would suggest a broadband power modulation, an interaction between *S* and *F* (low- vs. high-frequency bands) would lend support to a frequency-specific oscillatory correlate. Given prior findings of delta and theta band activity correlating with target-lure status and of gamma-band activity correlating with memory load, we hypothesized that an oscillatory correlate of summed similarity would appear as an interaction between *S* and *F* or as a three-way interaction between *S*, *F*, and *T* (time epoch: early vs. late portion of the retrieval period).

For lures, we found reliable ($p < .011$) neural correlates of S in all five ROIs; however, only the temporal cortex and medial temporal lobe (MTL) exhibited reliable effects for both targets and

Table 2

Numbers of Patients and Electrodes in Each of Five Anatomical Regions of Interest

Region of interest	Brodmann areas	No. patients	No. electrodes
Dorsolateral prefrontal cortex	9, 10, 46	9	95
Temporal cortex	20, 21, 22	16	413
Medial temporal lobe	28, 35, 36	14	126
Hippocampus		6	32
Parietal cortex	1, 2, 3, 5, 7, 39, 40	13	126

lures, and in both of those ROIs we found significant interactions between *S* and *F*. In temporal cortex we observed a significant $S \times F$ interaction for both targets and lures; in MTL we observed a significant $S \times F \times T$ interaction for both targets and lures (see the complete table in Appendix B). To further explore these interactions, we examined whether regression betas showed a monotonic increase or decrease with *S* across both target and lure trials. The frequency bands that show such monotonic increases or decreases are shown in Table 3.

In temporal cortex, lower frequency oscillatory power (4–14 Hz) decreases with *S* and higher frequency oscillatory power (24–90 Hz) increases with *S* for both targets and lures. In the MTL, 4–9 Hz theta and 28–128 Hz gamma power increase with *S* in the early epoch, but the low-frequency effect switches sign in the late time bin (second half of the retrieval interval), resulting in a three-way interaction between *S, F,* and *T*. Figure 3 shows that in temporal cortex, theta power decreases with summed similarity and gamma power increases in the late epoch (this change in direction of the relation produces the significant interaction term). In MTL, both theta and gamma power increase with summed similarity but do so in the early epoch. The 2–4 Hz delta band does not show a consistent increase or decrease with summed similarity across target and lure trials.

Results from our scalp EEG studies (Experiments 2 and 3) are shown in Figures 4 and 5. These figures summarize the regression results for the targets model and the lures model, respectively. The first row of each figure shows clusters of electrodes (see Method) that exhibit significant oscillatory correlates of *S* that are not modulated by frequency band, retrieval epoch (early vs. late), or item recency. The subsequent rows of topographical plots show clusters of electrodes that exhibit significant interactions between *S* and frequency range *F* (low vs. high); between *S* and time epoch (*T*); and between *S,* frequency range, and time window. Red-shaded clusters indicate significantly positive regression coefficients, and blue-shaded clusters indicate significantly negative regression coefficients.

The most consistent effect across the two scalp EEG studies and the targets and lures models was the finding of clusters of primarily frontal but also left parietal electrodes that exhibited a significant interaction between summed similarity and frequency range (see Figures 4 and 5). In electrodes for which this interaction was significant, the oscillatory summed-similarity effect differed between lower frequencies (delta, theta, and alpha) and higher frequencies (beta and gamma). Blue electrodes indicate that the summed-similarity effect was more positive for the lower frequency bands; red electrodes indicate the opposite. Whereas in both experiments we observed similar clusters that exhibited a stronger summed-similarity effect for the lower frequency bands, in Experiment 2 we also observed a cluster of central electrodes that exhibited a summed-similarity effect that was more positive for high frequencies. In Experiment 2 we further observed clusters of significant electrodes that exhibited significant modulation of broadband power with summed similarity.

In Experiment 2, a set of frontal electrodes also showed an interaction between *S* and time epoch (see Figure 5c), indicating that the summed-similarity effect varied between the early and late epoch (i.e., first half vs. second half of the retrieval period). In this condition, frontal electrodes exhibited a more positive summed-similarity effect for the early than for the late epoch. There was a small group of right parietal electrodes for lures that exhibited a triple interaction between *S,* frequency, and time (see Figures 4d and 5d). The interaction between *S* and frequency band in these electrodes indicates that the frequency-specific summed-similarity effect changes over the course of the trial. There was no significant effect of lag or RT on oscillatory activity during retrieval.

To further examine the frequency specificity of the summed-similarity effect, we computed normalized oscillatory power in the ROI defined by the summed similarity-by-frequency interaction in

Table 3

Sign of Slope of Oscillatory Power on Similarity Bins, Grouped by ROI and Time Bin

Epoch and ROI	Frequency band						
	Delta (2–4 Hz)	Theta (4–9 Hz)	Alpha (9–14 Hz)	Beta (14–28 Hz)	Low-gamma (28–48 Hz)	Mid-gamma (48–90 Hz)	High-gamma (90–128 Hz)
	Experiment 1						
Early							
TC	0	0	−	−	−	0	+
MTL	0	+	0	0	0	+	+
Late							
TC	0	−	−	0	+	+	+
MTL	0	0	0	0	0	+	+
	Experiment 2 (and Experiment 3)						
Early							
Frontal	0 (0)	+ (0)	+ (0)	0 (0)	+ (0)	+ (0)	0 (0)
Parietal	**+ (+)**	0 (0)	+ (0)	+ (0)	+ (0)	0 (0)	+ (0)
Late							
Frontal	0 (0)	− (+)	− (0)	0 (0)	0 (0)	0 (0)	0 (0)
Parietal	+ (0)	0 (+)	+ (0)	+ (0)	− (0)	0 (0)	0 (0)

Note. ROIs for Experiment 1 are based on Table 1; ROIs for Experiments 2 and 3 are based on Figures 4b and 5b. In the lower half of the table, the first symbol refers to Experiment 2, and the second symbol refers to Experiment 3. A plus sign denotes a positive slope for both targets and lures. A minus sign denotes a negative slope for both targets and lures. Boldface indicates frequency bands that have consistent slopes between targets and lures in both experiments. 0 = no consistent slopes between targets and lures; ROI = region of interest; TC = temporal cortex; MTL = medial temporal lobe.
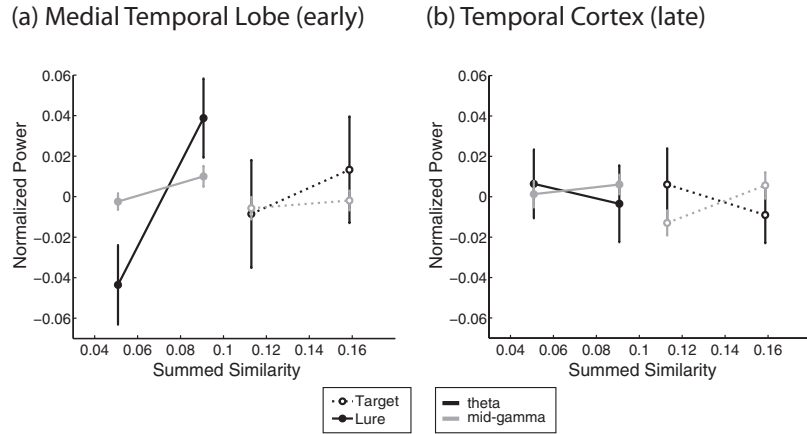
Figure 3. Experiment 1. Average normalized oscillatory power as a function of summed-similarity bin is shown (a) in the early time bin in the medial temporal lobe and (b) in the late time bin in temporal cortex. In the medial temporal lobe, both 4–9 Hz theta and 48–90 Hz gamma power increase with summed similarity. In temporal cortex, theta decreases whereas gamma increases with summed similarity. Error bars reflect standard error of the mean.

Figure 5b. The ROIs were kept the same between Experiments 2 and 3 by taking the intersection of electrodes that were significant for both experiments. We then examined which of the frequency bands (2–4 Hz delta, 4–9 Hz theta, 9–14 Hz alpha, 14–28 Hz beta, 28–48 Hz low gamma, 48–90 Hz mid gamma, and 90–128 Hz high gamma) displayed the pattern of regression slopes expected of a neural correlate of summed similarity. A neural correlate of summed similarity should show a correlation with summed similarity in the same direction for targets and lures. Because we fit our regression model separately for targets and lures, a significant summed similarity-by-frequency band interaction in the regression model does not guarantee that this condition is satisfied. Table 3
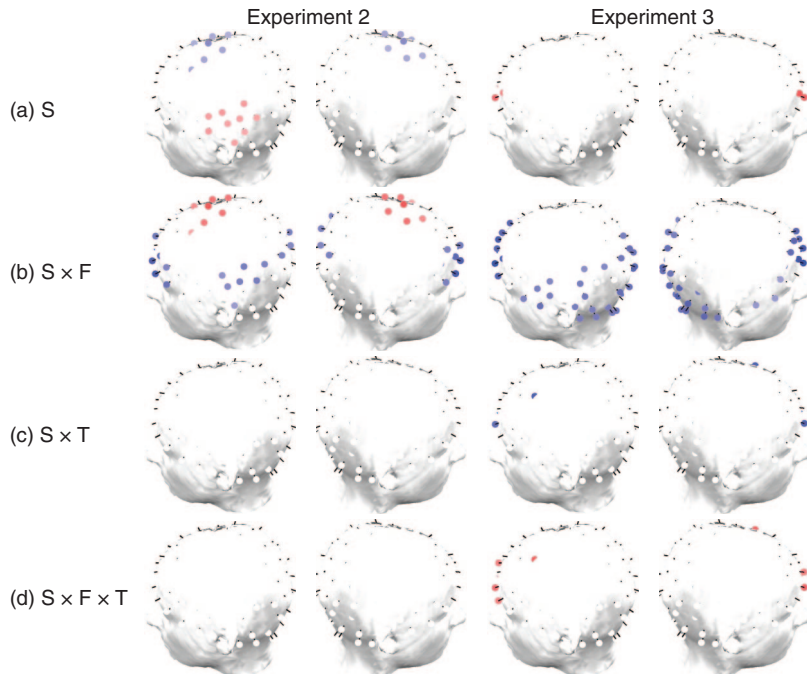

Figure 4. Experiments 2 and 3. Topographical plots of significant ($p < .05$) electrode clusters for regression coefficients of the targets model. Electrodes shown in red exhibit a significantly positive regression coefficient across participants for (a) summed similarity (S), (b) interaction between S and frequency (F), (c) interaction between S and time (T), (d) triple interaction between S, frequency, and time. Similarly, electrodes shown in blue exhibit a significantly negative regression coefficient across participants. We do not show topographical plots for lag or response time, because those regression coefficients produce no statistically significant clusters.
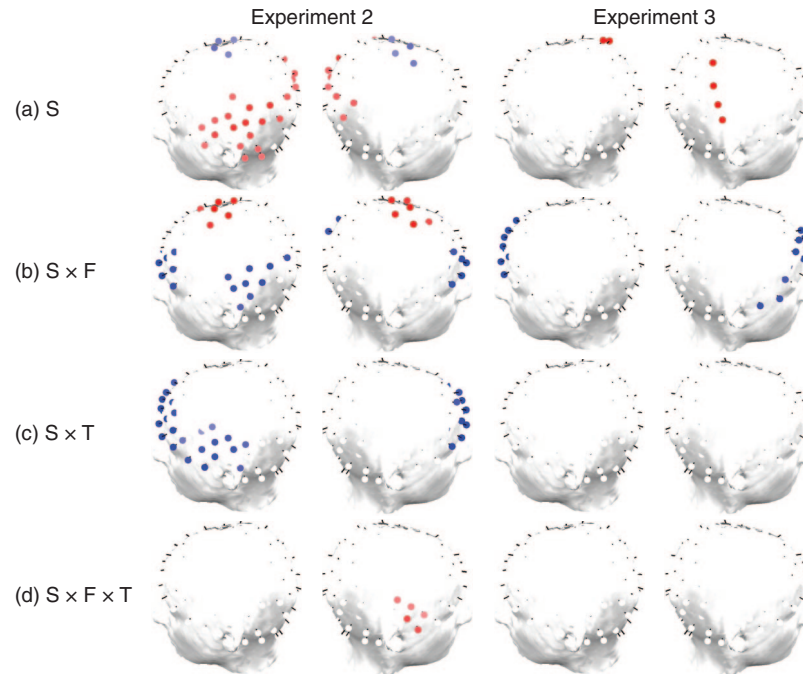
Figure 5. Experiments 2 and 3. Topographical plots of significant ($p < .05$) electrode clusters for regression coefficients of the lures model. Electrodes shown in red exhibit a significantly positive regression coefficient across participants for (a) summed similarity (S), (b) interaction between S and frequency (F), (c) interaction between S and time (T), (d) triple interaction between S, frequency, and time. Similarly, electrodes shown in blue exhibit a significantly negative regression coefficient across participants. We do not show topographical plots for lag or response time, because those regression coefficients produce no statistically significant clusters.

shows that there are patterns of oscillations in all frequency bands that satisfy this criterion.

In Experiment 3 we also observed a significant interaction between summed similarity and oscillatory frequency in frontal and left parietal electrodes (see Figures 4c and 5c). We did not observe the negative interaction between summed similarity and frequency in central electrodes that was present in Experiment 2. There was a significant interaction between S and time epoch in a few frontal electrodes. There was no significant triple interaction between S, time, and frequency band. The main effect of summed similarity occurred in a small set of right parietal and bilateral frontal electrodes, which increased in oscillatory power with S. As in Experiment 2, there was no statistically significant effect of lag or RT on oscillatory activity.

Considering the data from the two scalp EEG studies together, the most consistent positive correlate of summed similarity across both targets and lures and across both experiments is the delta band activity observed in parietal and frontal cortices (see Figure 6). The theta band shows a pattern similar to that in the delta band, but the direction of the theta effect is not consistent across the same time epochs in the two experiments (see Figure 7). In Experiment 2, theta increases with S in the first time bin, whereas in Experiment 3 it increases with S in the second time bin. Although the increase in delta activity with S replicated nicely across both scalp studies, the effect was somewhat less consistent in the theta band, appearing at different time epochs in Experiments 2 and 3. This difference is not too surprising, given the two major

differences in procedure across the experiments: In Experiment 3, participants studied a fixed set of lists many times, whereas in Experiment 2, they saw each list once. Another possibility is that the difference is driven by the fact that in Experiment 2, different list lengths were intermixed, which was not the case in Experiment 3. Consequently, participants had to continuously update their decision threshold in Experiment 2 but did not have to do so in Experiment 3.

General Discussion

A number of the most successful cognitive theories of recognition memory posit that people judge a test item as old when its summed similarity to the list items exceeds a threshold value (e.g., Hintzman, 1988; Kahana & Sekuler, 2002; Lamberts et al., 2003; Nosofsky, 1991; Nosofsky & Kantner, 2006; Shiffrin & Steyvers, 1997). This summed-similarity signal is regarded as a likely mechanism for the familiarity component of dual process theories of recognition (Rugg & Yonelinas, 2003). The same computational account has also been offered as the basis for categorization decisions when categories are defined by complex regions in multidimensional feature space (Fific, Little, & Nosofsky, 2010). In view of summed similarity's broad significance as a latent construct in cognitive theories of memory, we asked whether summed similarity exhibited a distinct electrophysiological signature beyond the electrophysiological correlates of recognition accuracy, study-test lag, and other experimental factors.
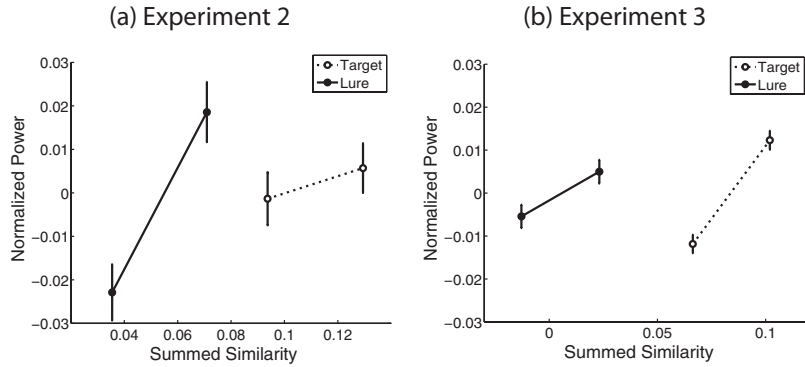
Figure 6. Neural summed-similarity effect for Experiments 2 and 3. Normalized oscillatory power shown as a function of binned summed similarity (defined in Equation 1) for both targets (dashed lines) and lures (solid lines). Bins are defined as the lowest and highest 50% of the summed-similarity distribution. Power is averaged over all frequencies in the 2–4 Hz delta band and in the channels within the frontal region of interest (see Figures 4b and 5b) in the first half of the trial. Error bars reflect standard error of the mean.

Across three experiments we found that brain oscillations measured using depth electrodes and cortical surface electrodes in temporal cortex and the medial temporal lobe (Experiment 1) and measured at the scalp (Experiments 2 and 3) are reliably modulated by summed similarity both for targets and for lures. In particular, we found that lower frequency oscillations, in the delta and theta bands, increased with summed similarity for both targets and lures in the medial temporal lobe (Experiment

1) and in clusters of parietal and frontal electrodes (Experiments 2 and 3). Although the general trend is present in all three studies, there were also differences between the experiments: Summed similarity correlated with theta band oscillations in the medial temporal lobe, whereas in the scalp recordings the effect was found in the delta band. Between the two scalp experiments, which varied in the list lengths used and the degree of practice that participants had, we observed the summed-
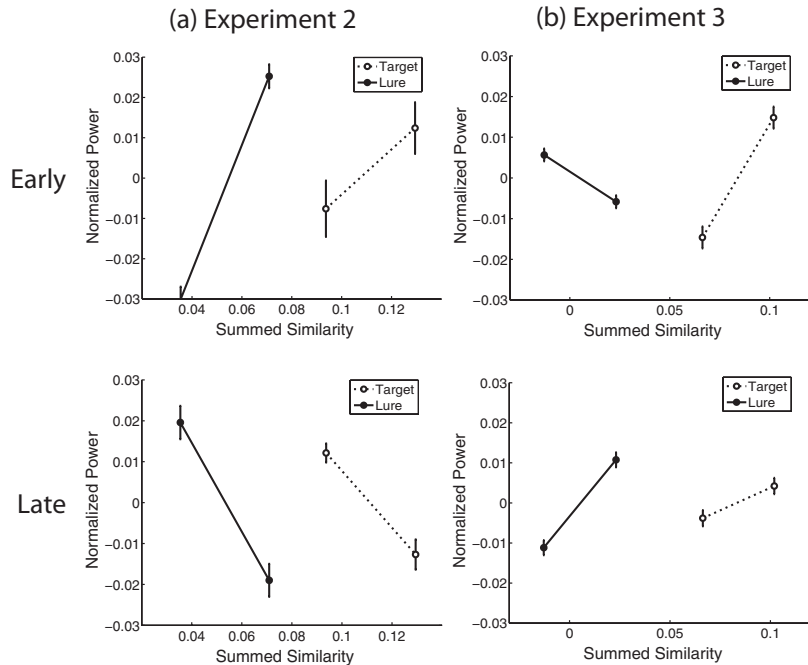

Figure 7. Neural summed-similarity effect in the 4–9 Hz theta band for Experiments 2 and 3. Normalized oscillatory power as a function of binned summed similarity (defined in Equation 1) for both targets (dashed lines) and lures (solid lines). Bins are defined as the lowest and highest 50% of the summed-similarity distribution. Power is averaged over all frequencies in the 4–9 Hz theta band and in the channels within the parietal region of interest (see Figures 4b and 5b), separately for the first half of the trials (Early) and the second half of the trials (Late). Error bars reflect standard error of the mean.

similarity correlate in the theta band earlier in the retrieval period in Experiment 2 and later in Experiment 3.

Although the prior literature suggests that theta oscillations are a stronger correlate of target-lure status than are delta oscillations, the strongest effects in the literature have often been observed at around 4 Hz, which is right at the boundary between the traditional theta and delta frequency bands. In addition, several studies have observed that delta and theta oscillations are similarly modulated during cognitive tasks (e.g., Jacobs et al., 2006; Lega, Jacobs, & Kahana, 2012; Mormann et al., 2005).

Related to the present work, several recent studies have begun to look at the neural correlates of similarity itself. For example, Loffler et al. (2005) showed that in the fusiform face area, the fMRI BOLD signal responds to changes in structural distance between a mean face and a face further away in the face space. Similarly, Aguirre showed that distance in similarity space was associated with the magnitude of fMRI adaptation in the fusiform face area (Aguirre, 2007). As such, they observed a neural similarity matrix that was correlated with the similarity matrix derived from the stimulus structure. Similarly, Kriegeskorte et al. (2008) demonstrated the existence of neural similarity matrices of a wide range of objects and faces in area IT (inferior temporal cortex), in both BOLD signal and single-unit activity. For verbal stimuli, Manning, Sperling, Sharan, Rosenberg, and Kahana (2012) showed that predominantly low-frequency oscillations across the cortex varied with their semantic similarity distances. Moving from perception to memory, Yago and Ishai (2006) compared the BOLD response in an item recognition task to probes that were the same as, ambiguous, similar to, or dissimilar to a studied prototype image (painting). They found an increase in BOLD activity with the level of similarity between probe and study item in the bilateral inferior parietal sulcus and superior parietal lobe. In addition, they found a decrease of BOLD activation with probe-item similarity in the superior frontal gyrus and precuneus.

A key assumption in summed-similarity theories of item recognition is that participants use all items in the contextually defined memory set to help distinguish between targets and lures. This assumption leads to the prediction that participants are more likely to endorse a target or a lure when there are other similar items stored in memory. However, summed-similarity theory also assumes a nonlinear (exponential) function relating similarity to distance in psychological space. The rate at which similarity decays with distance is a parameter of the model, and for the decay-rate parameters estimated in most studies, the actual contribution of most items to the summed-similarity computation will be negligible. This means that in practice a summed-similarity model will give most weight to items that are very similar to the target or lure and that less similar list items contribute minimally to the summed-similarity calculation. Because of this nonlinearity, it would require considerable statistical power to distinguish a summed-similarity model from a model that uses only the similarity of the most similar lure or target. Prior studies that have carefully explored this issue have shown that the summed-similarity models do in fact provide a better quantitative fit to behavioral data on item recognition than models considering only the most similar list item (e.g., Huang, Kahana, & Sekuler, 2012; Kahana et al., 2007).

Our study provides neural evidence for the summed-similarity computation hypothesized to underlie recognition memory and category learning (Brown, Neath, & Chater, 2007; Hintzman, 1988; Kahana & Sekuler, 2002; Lamberts et al., 2003; Nosofsky, 1991). Similarity-based interference is reflected in frontal and parietal delta and theta oscillations measured in scalp EEG and in theta oscillations recorded from depth electrodes in the human medial temporal lobe. Given that previous studies have found similar neural correlates for old–new item recognition and categorization (Nosofsky, Little, & James, 2012), we would expect to find similar oscillatory correlates of summed similarity in categorization tasks.

References

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. *NeuroImage, 35,* 1480–1494. doi:10.1016/j.neuroimage.2007.02.005

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57,* 289–300.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. doi:10.1037/0033-295X.114.3.539

Düzel, E., Neufang, M., & Heinze, H. J. (2005). The oscillatory dynamics of recognition memory and its relationship to event-related responses. *Cerebral Cortex, 15,* 1992–2002. doi:10.1093/cercor/bhi074

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review, 117,* 309–348. doi:10.1037/a0018526

Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods, 39,* 950–958. doi:10.3758/BF03192990

Guderian, S., & Düzel, E. (2005). Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans. *Hippocampus, 15,* 901–912. doi:10.1002/hipo.20125

Harris, A., & Aguirre, G. K. (2010). Neural tuning for face wholes and parts in human fusiform gyrus revealed by fMRI adaptation. *Journal of Neurophysiology, 104,* 336–345. doi:10.1152/jn.00626.2009

Heiser, W. J., & Bennani, M. (1997). Triadic distance models: Axiomatization and least squares representation. *Journal of Mathematical Psychology, 41,* 189–206. doi:10.1006/jmps.1997.1166

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in multiple-trace memory model. *Psychological Review, 95,* 528–551. doi:10.1037/0033-295X.95.4.528

Howard, M. W., Rizzuto, D. S., Caplan, J. C., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., . . . Kahana, M. J. (2003). Gamma oscillations correlate with working memory load in humans. *Cerebral Cortex, 13,* 1369–1374. doi:10.1093/cercor/bhg084

Huang, J., Kahana, M. J., & Sekuler, R. (2012). *Similarity effects in name–face recognition: A dual-process, summed-similarity account.* Manuscript submitted for publication.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage, 32,* 978–987. doi:10.1016/j.neuroimage.2006.02.018

Jacobs, J., & Kahana, M. J. (2010). Direct brain recordings fuel advances in cognitive electrophysiology. *Trends in Cognitive Sciences, 14,* 162–171. doi:10.1016/j.tics.2010.01.005

Kahana, M. J. (2012). *Foundations of human memory.* New York, NY: Oxford University Press.

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. doi:10.3758/BF03211679

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research, 42,* 2177–2192. doi:10.1016/S0042-6989(02)00118-9

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure-similarity affects visual episodic recognition: Detailed tests of a noisy examplar model. *Memory & Cognition, 35,* 1222–1232. doi:10.3758/BF03193596

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141. doi:10.1016/j.neuron.2008.10.043

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. doi:10.1037/0096-3445.132.3.351

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus, 22,* 748–761. doi:10.1002/hipo.20937

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8,* 1386–1391. doi:10.1038/nn1538

Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., & Kahana, M. J. (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *Journal of Neuroscience, 32,* 8871–8878.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190. doi:10.1016/j.jneumeth.2007.03.024

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10,* 465–501. doi:10.1016/0010-0285(78)90008-7

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C., & Fernández, G. (2005). Phase/amplitude reset and theta–gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus, 15,* 890–900. doi:10.1002/hipo.20117

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,* 3–27. doi:10.1037/0096-1523.17.1.3

Nosofsky, R. M., & Kantner, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition, 34,* 112–124. doi:10.3758/BF03193391

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review, 118,* 280–315. doi:10.1037/a0022494

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences, USA, 109,* 333–338. doi:10.1073/pnas.1111304109

Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences, 7,* 313–319. doi:10.1016/S1364-6613(03)00131-1

Shepard, R. N. (1987, September 11). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. doi:10.1126/science.3629243

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. doi:10.3758/BF03209391

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42,* 7–67. doi:10.1007/BF02293745

van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A., & Kahana, M. J. (2010). Hippocampal gamma oscillations increase with working memory load. *Journal of Neuroscience, 30,* 2694–2699. doi:10.1523/JNEUROSCI.0567-09.2010

van Vugt, M. K., Schulze-Bonhage, A., Sekuler, R., Litt, B., Brandt, A., Baltuch, G., & Kahana, M. J. (2009). Intracranial electroencephalography reveals two distinct similarity effects during item recognition. *Brain Research, 1299,* 33–44. doi:10.1016/j.brainres.2009.07.016

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. doi:10.1371/journal.pbio.0050056

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research, 42,* 2909–2923. doi:10.1016/S0042-6989(02)00362-0

Workman, W., & Adams, O. (1950). Suggestions for simplifying the computation of vincent curves. *Journal of General Psychology, 43,* 147–151. doi:10.1080/00221309.1950.9710613

Yago, E., & Ishai, A. (2006). Recognition memory is modulated by visual similarity. *NeuroImage, 31,* 807–817. doi:10.1016/j.neuroimage.2005.12.014

Yotsumoto, Y., Kahana, M. J., Wilson, H. R., & Sekuler, R. (2007). Recognition memory for realistic synthetic faces. *Memory & Cognition, 35,* 1233–1244. doi:10.3758/BF03193597

(*Appendices follow*)

Appendix A

Multidimensional Representation of Synthetic Faces

Faces were constructed from four orthogonal dimensions, as shown in Figure A1. Inspection of the figure and of the measurement vector that makes up each of the four theoretical dimensions demonstrates that the first orthogonal dimension corresponds mostly to head width and mouth height. The second orthogonal dimension corresponds mostly to a hairline parameter and lip thickness. The third dimension is related to the curvature of the bottom of the face, and the fourth dimension reflects curvature of the hairline.

The empirically derived MDS coordinates for this face set are shown in Table A1. They map onto slightly different sets of measurements. One of the reasons for this difference with the
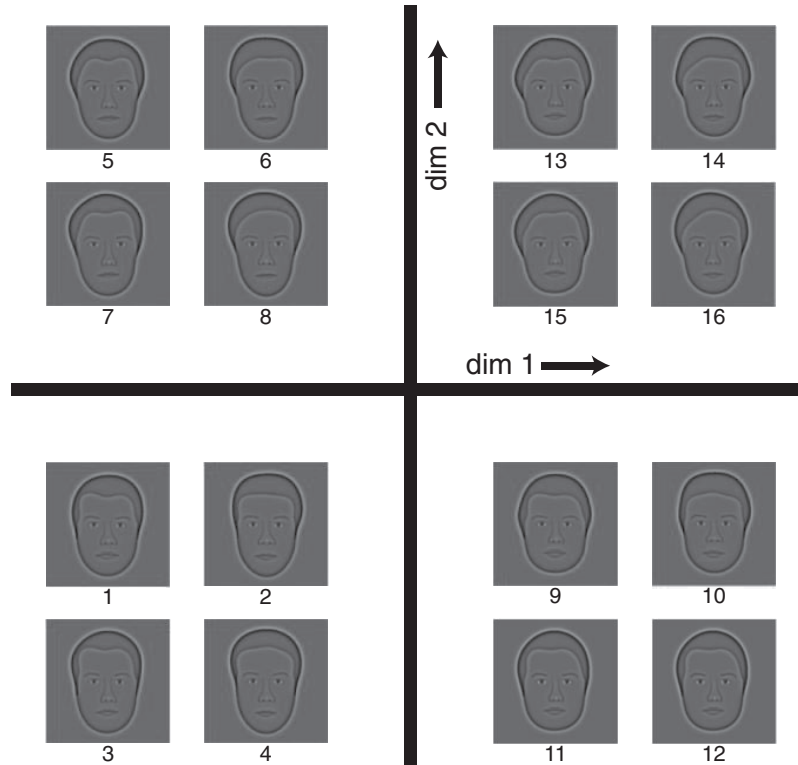

Figure A1. Two-dimensional projection of the four-dimensional theoretical face space. Each face was constructed by moving a fixed distance (distance-unit), equal to the 75% correct discrimination threshold (just noticeable difference), in positive or negative directions away from the mean face. All possible permutations of positive and negative shifts for the four principal components of face space yielded a set of 16 faces. Each quadrant shows four faces that have the same coordinates for dimension 1 and 2 but differ in their coordinates for dimensions 3 and 4. For example, the faces in the right bottom quadrant are constructed by a positive distance-unit in the direction of dimension 1 and a negative distance-unit in the direction of dimension 2. For each of the four faces, the coordinates in the third and fourth dimension are different. The numbers below each face correspond to those in Table A1. dim = dimension.

(*Appendices continue*)

Table A1

Similarity Coordinates of the Face Stimuli Derived From the MDS Study

Face	dim 1	dim 2	dim 3	dim 4
1	0.901	0.706	1.060	−1.079
2	1.886	−0.683	−0.227	0.498
3	1.128	0.942	0.022	−1.137
4	1.284	−1.257	−0.231	0.891
5	−0.075	1.511	−1.029	−0.483
6	0.484	−0.390	−1.059	1.690
7	0.384	1.504	−0.951	0.552
8	0.268	−0.468	−0.106	2.199
9	−0.723	0.215	1.953	0.039
10	−0.248	−1.602	0.534	−0.860
11	−0.342	0.087	1.801	−0.789
12	0.402	−1.269	−0.481	−1.391
13	−1.298	1.292	0.925	0.165
14	−1.249	−0.599	−1.242	−0.532
15	−1.725	0.840	0.301	−0.444
16	−1.079	−0.828	−1.271	0.682

Note. MDS = multidimensional scaling; dim = dimension.

theoretically derived coordinates is that the empirically derived MDS coordinates were not Procrustes transformed. We chose not to Procrustes transform the empirically derived MDS coordinates, such that we would not alter the data. The first dimension correlates highest with nose width, $r(15) = .89$, $p < .001$, and one of the components of the head shape, $r(15) = .88$, $p < .01$. The second dimension from the MDS study correlates highest with a hairline parameter, $r(15) = .89$, $p < .001$, as well as eye separation, $r(15) = .85$, $p < .001$. The third dimension correlates highest with thickness of the top lip, $r(15) = 0.82$, $p < .001$, as well as a component of head shape, $r(15) = −.73$, $p < .01$. Finally, MDS dimension four correlates most strongly with thickness of the top lip, $r(15) = −.72$, $p < .01$, and with head shape, $r(15) = 0.70$, $p < .01$.

This demonstrates that perception of these stimuli emphasizes specific aspects of the synthetic face stimuli (hairline, head shape, and lips), which led us to use the empirically derived MDS coordinates in our summed-similarity analyses. This was further justified by the fact that we found very similar attentional weights on these four dimensions (see Method).

Appendix B

Complete Overview of the Normalized Regression Betas for Experiment 1

	DLPFC		TC		MTL		Hippocampus		Parietal	
Model	Targets	Lures	Targets	Lures	Targets	Lures	Targets	Lures	Targets	Lures
S	0.021	0.044	0.008	0.041[*]	0.011	0.044[*]	0.032	0.060	0.022	0.032
RT	−0.003	0.007	−0.009	−0.005	−0.015	0.002	−0.013[*]	0.014	−0.012	0.002
S × F	−0.018	0.044[*]	**−0.023[*]**	**−0.049[*]**	−0.003	−0.023	−0.032	−0.040	−0.010	−0.025
S × T	−0.017	−0.060[*]	−0.008	−0.044[*]	**−0.025[*]**	**−0.068[*]**	−0.005	−0.038[*]	−0.015	−0.060[*]
S × F × T	0.023	0.069[*]	0.007	0.041[*]	**0.017[*]**	**0.055[*]**	0.003	0.041	0.016	0.066[*]

Note. Rows are regression coefficients (S = summed similarity; F = frequency band; T = time bin (early vs. late portion of the retrieval period); RT = response time; × = interaction term). Columns are brain regions (DLPFC = dorsolateral prefrontal cortex; TC = temporal cortex; MTL = medial temporal lobe). Models are T (targets model) and L (lures model). Boldface type denotes significant for both targets and lures. An asterisk denotes significance with a 0.05 false discovery rate (corresponding to a *p*-value threshold of .011).