.

Similarity effects in name-face recognition:
A dual-process, summed-similarity account

Jie Huang, Michael J. Kahana* & Robert Sekuler
Volen Center for Complex Systems and Department of Psychology
Brandeis University, Waltham MA
*Department of Psychology, University of Pennsylvania, Philadelphia, PA

We examined how similarity influences the ability to associate names and faces. In Experiment 1, multidimensional scaling generated perceptual similarity spaces for novel stimuli: synthetic human faces and single-syllable Mandarin Chinese names. In Experiment 2's associative recognition task, face-name pairs were presented in alternating Study and Test blocks. In every Study block, subjects were given the same 18 pairs of faces and names; in each Test block, face-name pairs that had been studied were intermixed with face-name pairs that had been rearranged. Subjects judged whether a face-name pair was one that had been studied, or was a new, rearranged pairing. With repeated study opportunities, associative recognition improved as both correct endorsements of studied pairs and correct rejections of rearranged pairs increased. Results were well fit by NEMo.D, a model in which a test probe's summed similarity to the contents of memory is supplemented by a similarity-dependent recollection process. NEMo.D's relationship to other models for recognition is explored.

Keywords: Recognition, learning, association, names, faces, Chinese, dual process model

The ability to associate names and faces is one of our most important social skills, and like other kinds of associations, stimulus similarity is an important determinant of the ease with which face-name associations are encoded and retrieved (Bower, Thompson-Schill, & Tulving, 1994; Nelson, Bajo, McEvoy, & Schreiber, 1989). For example, when a teacher must learn students' names, the task becomes particularly daunting when many look alike, or when their names sound similar, or both (e.g., Fraas et al., 2002).

In a study of similarity effects in memory for name-face associations, Pantelis, Vugt, Sekuler, Wilson, and Kahana (2008) asked subjects to learn associations between distinctive, common names, and visually confusable synthetic faces whose similarities were parameterized and manipulated in a four-dimensional feature space. They found that recall performance diminished as a function of the number of other studied faces that were perceptually similar to the probe face, and that incorrect recalled names were more likely to be names associated with faces that were visually similar to the probe face. A similar pattern of results was obtained in an associative recognition task where subjects first studied a series of name-face pairs and later

judged whether or not a test pair matched one of the studied pairs: subjects tendency to endorse a target pair decreased as the number of similar faces in the study list increased.

In the present study we further examine similarity effects in associative recognition by manipulating both the similarity among faces and the similarity among names. Our key question is whether summed-similarity models that have been highly successful in accounting for similarity effects in item recognition tasks (e.g., Nosofsky, Little, Donkin, & Fific, 2011) can also account for the observed pattern of similarity effects in associative recognition. Because a given name (or face) can also directly retrieve (recall) its associated face (or name) from memory, we also explored a model in which a similarity-based familiarity signal is combined with a similarity-based recall (or recollection) process. Of particular interest was whether a dual-process similarity-based model of associative recognition could substantially outperform a single process similarity-based model.

According to summed-similarity models (e.g., Kahana & Sekuler, 2002; Nosofsky et al., 2011), recognition decisions are based on the sum of the pairwise similarities between the probe item and each of the items stored in a contextually-defined memory array (usually the items on a just-presented list). Conceptually one can think of each item representation as resonating to some degree with the probe, with the sum of the resonances providing a familiarity signal upon which recognition decisions are based. In this way, a lure item may be endorsed as a target when it is highly similar to multiple items in the study list, as would be the case in experimental paradigms designed to induce false memories (e.g., Deese, 1959; Roediger & McDermott,

1995).

Our study began by identifying suitable stimuli whose similarity relationships could be measured and leveraged to theoretical advantage. Additionally, in order to track the course of learning, we chose stimuli that would initially be unfamiliar to test subjects. So, for name stimuli we generated a set of spoken single-syllable Mandarin Chinese names. As our subjects had no experience with Asian languages, we could be certain that these stimuli would be unfamiliar. The use of these unfamiliar names also negated the confounding influence of pre-experimental facial stereotypes associated with some familiar names (Lea, Thomas, Lamkin, & Bell, 2007). Stimulus faces were drawn from a library of realistic, synthetic human faces. Various subsets of these "Wilson faces" (named for the eponymous H. R. Wilson) have been used in studies of face perception, face memory, and facial expression of emotion (e.g., Wilson, Loffler, & Wilkinson, 2002; Loffler, Yourganov, Wilkinson, & Wilson, 2005; Isaacowitz, Wadlinger, Goren, & Wilson, 2006; Yotsumoto, Kahana, Wilson, & Sekuler, 2007; Galster, Kahana, Wilson, & Sekuler, 2009; Pantelis et al., 2008). Over repeated study-test trials we asked subjects to learn arbitrary pairings of novel faces *and* novel names and tested their memory for those pairings using an associative recognition procedure. Our first goal was to define how similarity of names and faces each influenced associative recognition. Our second goal was to determine whether a single-process summed similarity model of perceptual recognition memory could account for these effects. Finally, we introduce a new dual-process model of associative recognition in which familiarity is modeled as a summed-similarity signal, while recollection is modeled as a similarity-dependent recall process. By comparing the success of single and dual process models we sought to shed further light on the ongoing debate concerning the contribution of recollective processes to associative recognition (Malmberg, 2008).

Experiment One

Experiment One used multidimensional scaling (MDS) to generate separate similarity spaces for the faces and names that would be used in Experiment Two. The similarity judgments needed for MDS were collected by the method of triads (Torgerson, 1958; Yotsumoto et al., 2007). On each trial, subjects were presented with three stimuli, and had first to identify the two that perceptually were most similar, and then the two that were most different. Then, the set of judgments was transformed offline into a representation of the faces' and names' perceptual spaces. The trio of stimuli presented on each trial was governed by an efficient, balanced incomplete block design (Weller & Romney, 1988).

Subjects

Eight subjects (one male; ages 19 to 21), who were naive to the purpose of the experiment, were paid for their participation. All had normal or corrected-to-normal vision as measured with Snellen targets. None of them reported problems with face perception, nor any experiences with an Asian language.

Apparatus

Both experiments were run under control of MATLAB scripts, which were supplemented using the Psychophysics Toolbox (Brainard, 1997). Face stimuli, generated in MATLAB, were presented on a 14-inch computer monitor set to refresh at 95 Hz; screen resolution was 1152 by 864 pixels. The display's mean luminance was maintained at 32 cd/m^2. During testing, a subject sat with head supported by a chin rest, viewing the computer display binocularly from 57 cm. Stimulus names were were pre-recorded and then presented through audio speakers.

Stimuli

Faces. Our face stimuli were 19 realistic synthetic faces[1] derived from photographs of 19 different Caucasian females and males. The methods used to transform the photographs to the stimuli that we use were described elsewhere (Wilson et al., 2002). In summary, the synthetic faces were low-pass filtered at 10 cycles per face width, an optimal value for face processing. In addition, this filtering neutralized distinguishing attributes such as hair color or curliness, and skin texture. All of our stimulus faces were shown in frontal view. Figure 1 illustrates the set of 19 faces (10 female and 9 male faces) that we used to generate the MDS-based similarity space for faces.

Names. The stimulus set comprised 19 Mandarin Chinese names that native speakers would perceive as being monosyllabic. The names are given in Table 1.[2] For brevity, we refer to Mandarin Chinese as "Chinese." Restricting the stimulus set to one-syllable names eliminated number of syllables as an extraneous factor that could influence memorability. All name stimuli possessed what linguists call the "second tone" (a.k.a., the rising tone) –a sound that rises from mid-level pitch to high. We had two reasons for choosing spoken Chinese names as stimuli. First, that choice should minimize pre-existing associations that familiar names might carry, and which might vary from subject to subject. Also, using Chinese names that are unfamiliar to Asian-language naive subjects should encourage subjects to base their similarity judgments on the

---

[1] A Balanced Incomplete Block design (BIBD) (Weller & Romney, 1988) controlled the number of trials needed for multidimensional scaling. The design of Experiment Two required 18, not 19 face-name pairs. However, no BIBD can be constructed for 18 items, although one can be constructed for 19 (Fisher & Yates, 1963). To work around this problem, the faces used for MDS included one "filler" stimulus of each class.

[2] Pinyin is the Romanization system for standard Mandarin. It is used to teach standard pronunciation, and to guide the spelling of Chinese names in non-Chinese publications. Wikipedia (http://en.wikipedia.org/Pinyin) explains that the pronunciation of a Chinese syllable is given in Pinyin by an initial (sound) that is followed by a final (sound), which is a combination of vowels.
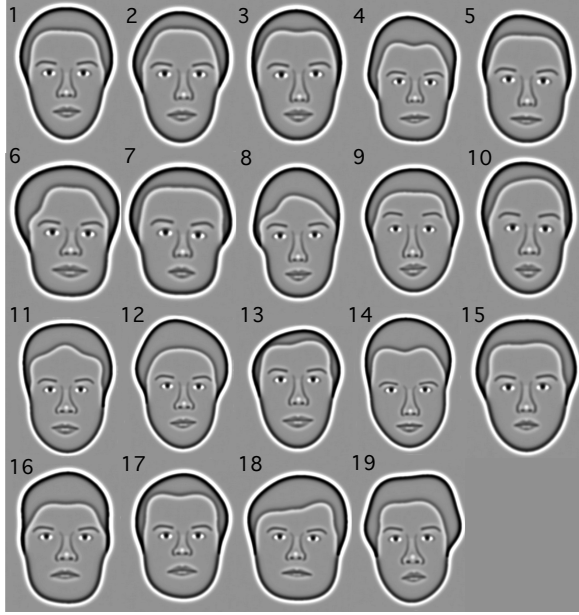
Figure 1. Faces used in Experiment One's multidimensional scaling. The stimuli in the upper two rows are female faces, those in the lower two rows are male faces.

Table 1
Nineteen Chinese names used in Experiment One. Names are given in Pinyin, the most-common Romanization system for Chinese. Also shown are the corresponding representations in the International Phonetic Alphabet (IPA), a system of phonetic notation based on the Latin alphabet. IPA is meant to represent the sounds of spoken languages in a standardized form. An audio presentation of the spoken names can be accessed at http://www.brandeis.edu/∼sekuler/mandarinNames.html.

Name ID	Pinyin	IPA	Name ID	Pinyin	IPA
1	lai	[laɪ]	11	qian	[tɕʰiɛn]
2	lan	[lan]	12	qu	[tɕʰy]
3	lou	[lɤʊ]	13	xia	[ɕia]
4	luo	[luɔ]	14	xue	[ɕyɛ]
5	lei	[leɪ]	15	xi	[ɕɨ]
6	qun	[tɕʰyn]	16	xiong	[ɕyʊŋ]
7	qie	[tɕʰiɛ]	17	cha	[tʂʰa]
8	qin	[tɕʰm]	18	chen	[tʂʰən]
9	quan	[tɕʰyɛn]	19	chuan	[tʂʰyɛn]
10	qiu	[tɕʰiou]			

phonological relationships among stimuli. The name stimuli were presented as audio recordings made by a female native speaker of Chinese (author J.H.). Her pronunciation of each name was captured in a 700 msec sound file.

Procedure

Face and name similarities were measured separately, using the method of triads (Torgerson, 1958; Yotsumoto et al., 2007). To control the number of trials needed for MDS, the number of times each pair of stimuli was presented was governed by a Balanced Incomplete Block design (BIBD) (Weller & Romney, 1988). With 19 stimuli, there are 171 unique pairs. BIBD makes it possible to present each possible pair once, distributed over 57 different triads (Fisher & Yates, 1963). Therefore, each subject served in 580 trials for the face stimuli and 580 for name stimuli. Face and name stimuli were presented in separate sessions. The first 10 trials in each session were treated as practice and were excluded from data analysis. The remaining 570 trials allowed each pair of faces and each pair of names to be presented 10 times. For each subject, trials within a block were presented in a new random order.

Multidimensional scaling of faces. To characterize faces' pairwise similarities, a triplet of faces was presented on each trial. A schematic of one trial is shown in Figure 2. The three faces were displayed simultaneously on a grey background, with each face subtending $6.80° \times 5.11°$ visual angle. The three faces were equidistant from a fixation point at the center of the computer display. At the viewing distance, 57 cm, the center of each face was $5.91°$ from fixation. Each trio of faces remained visible until the subject's responses were

made. To minimize the chance that a stimulus' location on the display might somehow influence subjects' responses, the spatial ordering of items in a triplet was randomized over trials.

Subjects made two responses on each trial. They first selected the two faces that were most similar, and then selected the two faces that were most different from one another. As a reminder of which judgment was to be made, the word "similar" in blue or the word "different" in red was presented at the center of the display. Subjects indicated their judgments by moving the computer cursor to the desired face, and then clicking the computer mouse. After a face had been selected, a color frame surrounded that face, confirming the selection. The color frame was blue during judgments of similarity, and red during judgments of difference. During any trial, subjects could correct an error by re-clicking on a face, thereby erasing its selection. Subjects were instructed not to base judgments on specific facial details (*e.g.*, eyes or nose), and they had unlimited time to make their responses.

Multidimensional scaling of names. It would confused subjects had multiple spoken name overlapped in time. Therefore, the sound files for each trial's triplet of names were played sequentially, with a 700 msec inter-stimulus interval. As a name was played, one of three squares on the computer display illuminated. The squares, labelled "A", "B", and "C", were spatially arranged on the display with the same geometry as for face stimuli. Before making similarity judgments, subjects could replay any or all of the names by clicking on the appropriate square. Replays were limited to
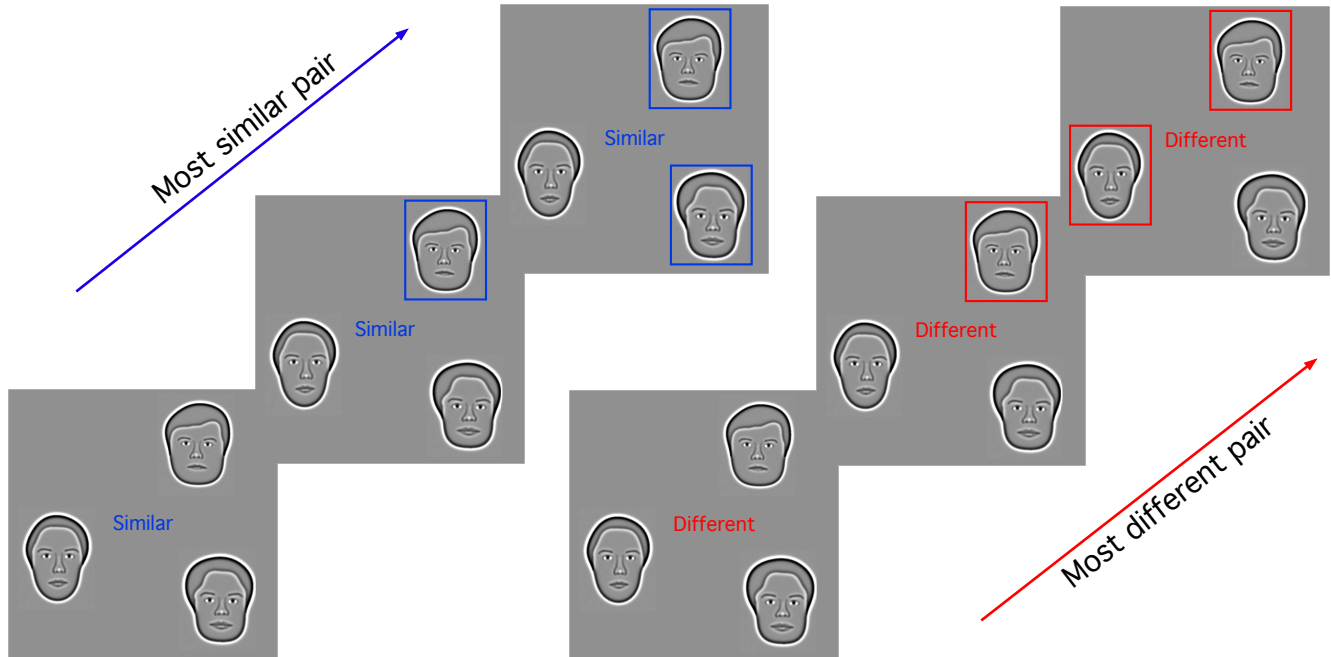
Figure 2.   Illustration of a trial used to estimate pairwise similarity within the set of faces in Experiment One. On each trial, subjects gave two judgments, first selecting the two faces that were most similar, as shown on the left side of figure, then selecting the two face that were most different, as shown on the right. Triplets of faces stayed on the display and were visible to subjects until both responses were made.

two per name per type of judgment ("similar" or "different"). The replays and the trial's consistent relationship between a spoken name and a square on the display allowed subjects to have the names firmly in memory. In making similarity judgments, subjects used the computer mouse to click on squares labelled "A", "B" or "C" identifying first the most two similar spoken names, and then the two most different names. Responses were made in the same way as on trials with face stimuli.

Results and Discussion

Subjects' responses were entered into 19×19 similarity matrices, one for faces and one for names. When a pair of stimuli had been judged most similar, a value of 3 was entered into the cell defined by the two stimuli; when a pair had been judged most different, a value of 1 was entered into the cells corresponding to the row and column for those stimuli. The undesignated stimulus pair earned the value 2 (Galster et al., 2009). For individual subjects, the pairwise similarity value in each matrix was normalized by the frequency with which a pair of stimuli had been presented. Using a weighted Euclidean model with SPSS' PROXSCAL routine, the normalized response matrices were transformed into similarity spaces by means of individual differences multidimensional scaling. MDS was carried out separately on judgments of names and for judgments of faces.

MDS for face stimuli MDS solutions for the face similarity space were generated in varying numbers of dimensions. As Figure 3A shows, Stress (Borg & Groenen, 2005) declined sharply as the number of dimensions increased from one to three, but saturated thereafter. In the three dimensional solution, Stress was 0.05 (the corresponding proportion of Dispersion Accounted For (DAF) was 0.95), showing that a three dimensional MDS configuration gives quite an adequate account of face similarity judgements. Figure 3B shows the Shepard plot associated with the three dimensional solution. There, the 171 inter-point distances derived from the three dimensional space are plotted against the corresponding pairwise similarity scores averaged across the eight subjects' judgments. The linear decrease in scatter shows that faces that were judged similar tended to lie close to one another in the MDS similarity space. In Figure 3C, each point represents the weights that an individual subject assigns to the three dimensions. As can be seen, the eight subjects clustered within a relatively small region, indicating that the subjects were in good agreement about the weights they gave to each dimension. The Stress values for the eight subjects were also tightly clustered around the average Stress value 0.05 (SD = 0.009, 95% CI [0.043, 0.058]). The disaggregated Stress values for the 19 faces also clustered tightly (SD = 0.006, 95% CI [0.0478, 0.0537]). Together, these results reveal no distinctly idiosyncratic faces in the stimulus set, and no idiosyncratic judgement strategies from particular subjects, which qualified the face stimuli and their
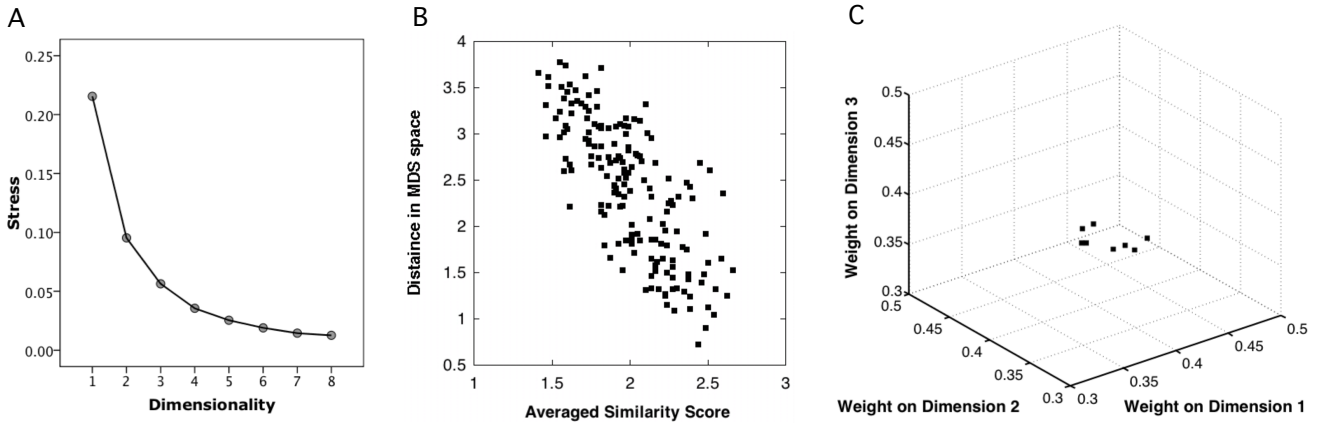
MDS of faces

Figure 3. Evaluating the MDS solutions for face stimuli. A: Stress values for MDS solutions in varying numbers of dimensions. Stress, which is inversely related to goodness of fit, declined substantially from a solution in one dimensional space to a solution in three dimensions. B: A Shepard plot of inter-point distances between faces in the three dimensional space versus pairwise similarity scores from subjects' judgments. Each point represents one face pair of the 171 unique pairs of faces. C: Individual subjects' weights on the three dimensions in the MDS solution; each point represents a single subject.

similarity space as suitable for use in Experiment Two, with other subjects.

Figure S1 shows the 19 faces' locations in the three dimensional similarity space. Although the interpretation of the dimensions is not consequential for our purposes, Dimension 1 seems to capture variation in head and chin shape, with a more positive value signifying that a chin and head that are wider and more square-shaped; Dimension 2 corresponds to the width and length of nose, and hair volume, with a more positive value signifying a wide nose and more hair; and Dimension 3 corresponds to the width of mouth and thickness of upper lip, with a more positive value being related to a narrower mouth with thicker lips.

MDS for name stimuli An analysis that paralleled the one just described for faces was used to generate a similarity space for name stimuli. As with faces, a three dimensional solution provided a very good fit to the similarities among the names (see Figure 4). Decomposed Stress values for individual subjects (ranging from 0.037–0.055), and for individual names (ranging from 0.0230–0.069), show no obvious outlier values for particular subjects or for particular names.

Figure S2 shows the 19 names' locations in the three dimensional similarity space. Visual inspection of the inter-item separation among names along each dimension suggests that the three dimensions capture the phonological combination of the initial and the component "i" in the final (Dimension 1); the phonological characteristics of names' initial (Dimension 2); and the phonological characteristics of the final, particularly the component of "n" (Dimension 3). The three-dimensional coordinates of the names will be used to compute the pairwise Euclidean distances needed to choose stimuli for and model results from Experiment Two.

Experiment Two

In this experiment we used the similarity spaces for names and faces as embodied in Experiment 1 to examine how similarity would influence associative recognition of arbitrary pairings of novel names and novel faces.

Subjects

Thirty subjects (7 male, ages 19-27), all naive to the experiment's purposes, participated for payment. To control previous experience and familiarity with the faces and names that would be used in Experiment Two, subjects who served in Experiment One were excluded, as were subjects who had prior experience with any Asian language, or a self-reported a problem with face perception. All had normal or corrected-to-normal vision as measured with Snellen acuity targets.

Stimuli

Our experimental design crossed three groups of faces and three groups of names, with groups having low, medium or high values of summed similarity. Each of the three summed-similarity groups comprised six faces or six names. The MDS spaces generated in Experiment One yield pairwise inter-item distances, rather than inter-item similarity *per se*. Although later, in modeling the resulting data, we use Shepard's exponential function to relate similarity to distance, our nine stimulus groups were based on inter-item distances. Given stimulus i, its summed distance was calculated as $\sum d_{i,j}$ $(j = 1, 2...19, \; j \neq i)$. In doing so, we treat the summed pairwise Euclidean *distances* in MDS space between any item and all other items in its stimulus class as a monotone inverse surrogate for similarity. As our design required just 18 name and 18 face stimuli, we
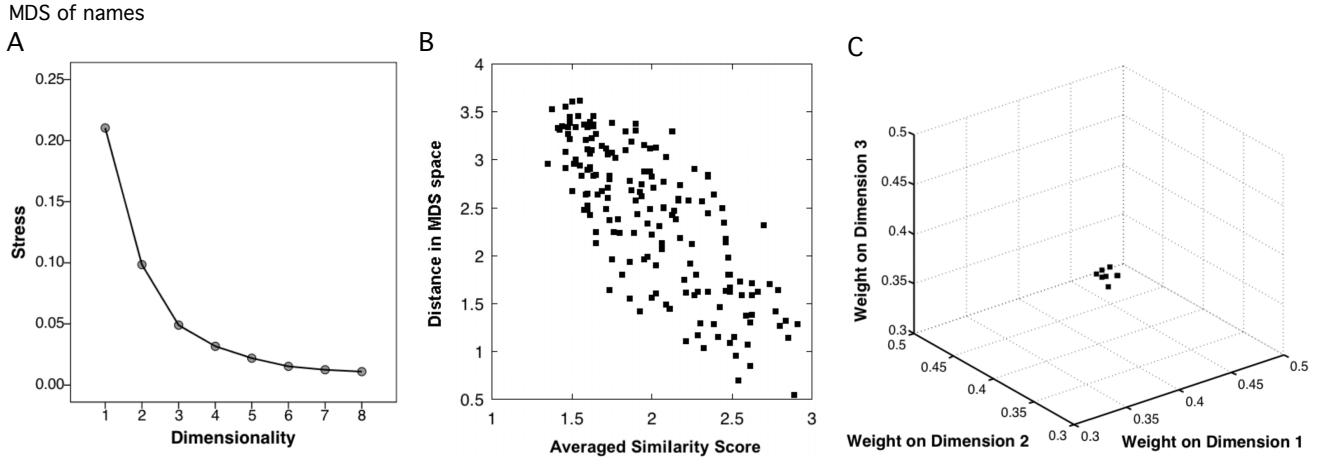
MDS of names

Figure 4. MDS solutions for 19 name stimuli. A: Stress values for MDS solutions for representations from one to eight dimensions. Similarly, stress value declined dramatically from one dimension to three dimensions, and was relatively constant thereafter. B: A Shepard plot of names' inter-point distance within the three dimensional solution space against pairwise similarity score transformed from subjects' judgments for names in Experiment One. Each point represents one name pair of the total 171 pairs ($\frac{19 \times 18}{2}$). C: Eight subjects' weights on the three derived dimensions in the name space. Note that each point represents a single subject.

dropped one name and one face from our original set of 19 stimuli. Specifically, we dropped face F13 and name N15 as each had a summed similarity value very close to that of another item, which made F13 and N15 redundant. Table 2 shows the summed-distance statistics for each group.

To generate the pairs of faces and names required by our experimental design, we constructed nine categories of face-name pairs by factorially combining three levels of summed similarity for faces with three levels of summed similarity for names (each category included two different exemplars, face-name pairs that shared the same levels of face summed similarity and name summed similarity). All subjects studied and were tested on the same sets of face-name pairs.

As we were interested in how an item's similarity to other items of the same class (faces or names) influenced associative recognition, we used Equation 1, which relates the similarity (η) of two stimuli to an exponential function of the Euclidean distance (*d*) between those items in the similarity space.

$$(1) \qquad\qquad \eta = ge^{-\tau d}$$

Procedure

The second experiment was carried out in two parts. First, subjects were familiarized with the synthetic face stimuli and with Chinese name stimuli, and demonstrated how readily the stimuli could be discriminated from one another; second, subjects took part in an associative recognition experiment. Both these parts are described below.

Same-Different Task. On each trial, subjects saw a pair of synthetic faces, or heard a pair of spoken Chinese names. After each pair was presented, subjects judged the items in the pair as "same" or "different." Figure 5A shows schematically the structure of a trial on which

"same"-"different" judgments for faces were collected. Each trial began with a fixation point presented for 500 msec at the center of the display. After a 500–600 msec uniform gray screen, a face was presented at the screen's center for 700 msec. This was followed by 700–800 msec during which the screen was again uniform gray. Then, a second face was presented for 700 msec, at the center of the screen. Subjects judged whether the two faces on a trial were the same or different. A similar procedure was adopted with name stimuli, except a name was presented by playing a 700-msec long sound file. Successive names on a trial were separated by 700–800 msec of silence. For both types of stimuli, on half the trials, the two stimuli were identical; on the other half, one stimulus was accompanied by a randomly-chosen item of the same class, face or name. Subjects received immediate feedback following a response. Trials with names and trials with faces were run in separate blocks, yielding a total of 72 trials per subject and stimulus type. Each stimulus within a stimulus class was presented equally often, allowing a subject to see or hear each stimulus eight times ($\frac{72 \times 2}{18}$) throughout this portion of the experiment.

Associative Recognition Task. The associative recognition task consisted of 10 alternating Study and Test blocks, with one minute between successive blocks. In each Study block, 18 face-name pairs were presented in random order (randomized separately for each subject and each block). For each pair, the face was presented at the center of the display for five seconds. The sound file containing the spoken name began to play 500 msec after the face's onset (see Figure 5B). Successive stimulus pairs were separated by a 1000 msec inter-trial interval; subjects were instructed to study and remember each pair.

In each Test block, 18 test pairs of faces and names were presented, but 9 of the pairs (*Targets*) preserved a pairing that had been studied, and 9 (*Lures*) rearranged studied items.

Table 2
Summed pairwise similarity for stimuli in categories of low, medium and high summed similarity.

	Stimulus identification number	Mean	SD	Minimum	Maximum
Faces					
Low Summed Similarity	F9, F18, F16, F14, F6, F7	46.13	1.48	44.68	48.54
Medium Summed Similarity	F4, F12, F10, F8, F3, F5	43.14	0.48	42.74	44.06
High Summed Similarity	F19, F15, F2, F11, F1, F17	40.72	1.08	39.13	42.39
Names					
Low Summed Similarity	N4, N3, N5, N2, N8, N1	45.59	0.86	44.85	46.68
Medium Summed Similarity	N7, N18, N11, N19, N13, N10	43.24	0.54	42.75	43.92
High Summed Similarity	N9, N17, N6, N16, N12, N14	40.90	0.67	39.75	41.78

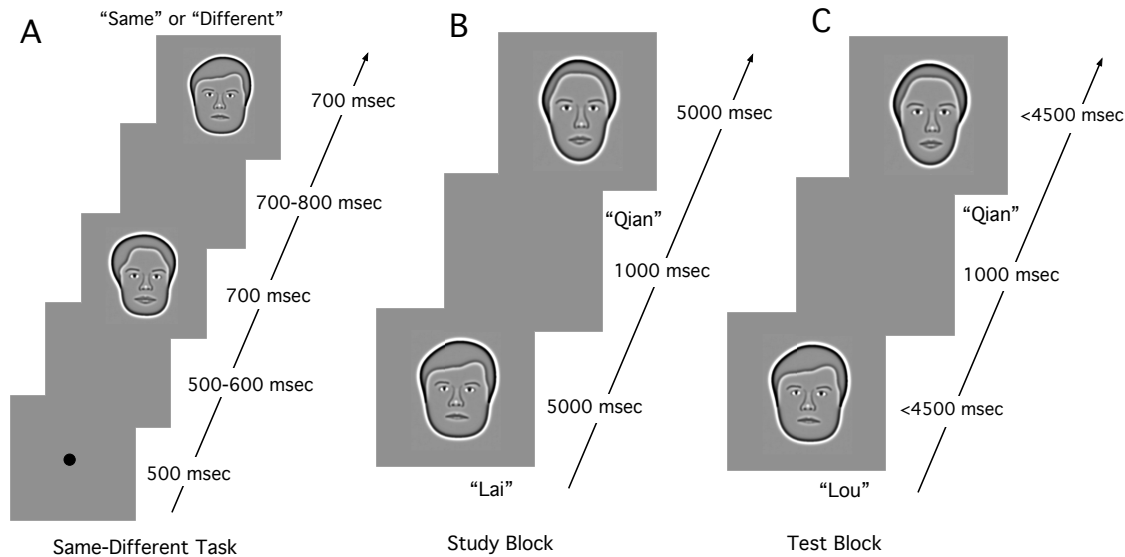*Figure 5.* Illustration of trial structure in the Same-Different task (A) and in the associative recognition experiment (B & C). A: Same-Different task for face stimuli. After seeing two sequentially-presented faces, a subject made a same-different judgment. Names were tested in a similar way, except that names were presented as pre-recorded sound files. B: In the Study blocks of the associative recognition experiment, the face in a pair was presented for 5000 msec, and its paired name (here, "Lai" or "Qian") was presented starting at 500 msec after the face's onset. A subject studied the stimuli and tried to remember them. C: In Test blocks of the associative recognition experiment, subjects were presented with face-name pairs and judged whether the pair was one that had been studied (compare the second stimulus shown here to its counterpart in B), or was a novel, re-arranged pairing (compare the first stimulus shown here to its counterpart in B). As explained in the text, after each face-name pair was presented, the subject's judgment was communicated using a 6-point rating scale.

Lure and *Target* pairs were presented in random order. The 9 intact pairs comprised 1 pair at each combination of the three levels of face and name summed similarity (see *Stimuli*). The 9 rearranged pairs were chosen such that the face and the name both came from the same level of summed similarity as the target. The items comprising an intact pair on block i could be part of two rearranged pairs on block $i + 1$. To mitigate learning of rearranged pairs, in such pairs no face was paired with any particular name more than once. To facilitate data analysis, in each Test block, subjects were all tested on the same set of preserved and rearranged pairs, though the pairs were presented in an order randomized anew for each subject.

The timing of face-name presentations was similar to the one used in Study blocks, except that the presentation of a face was limited to 4.5 seconds or when the subject responded, whichever came first. Subjects almost always responded before the time limit was reached (see Figure 5C). Subjects used a 6-point scale to report their confidence that the just-presented face-name pair was among the pairs that had been studied. Ratings ranged from "1" (sure "*No, this is not a studied pair*") to "6" (sure "*Yes, this is a studied pair*"). Intermediate ratings indicated more tentative judgments of whether the face and name had been studied as a pair (Macmillan & Creelman, 2005). Responses were communicated by pressing number keys on a computer's keyboard. A new trial began one second after the response was recorded.
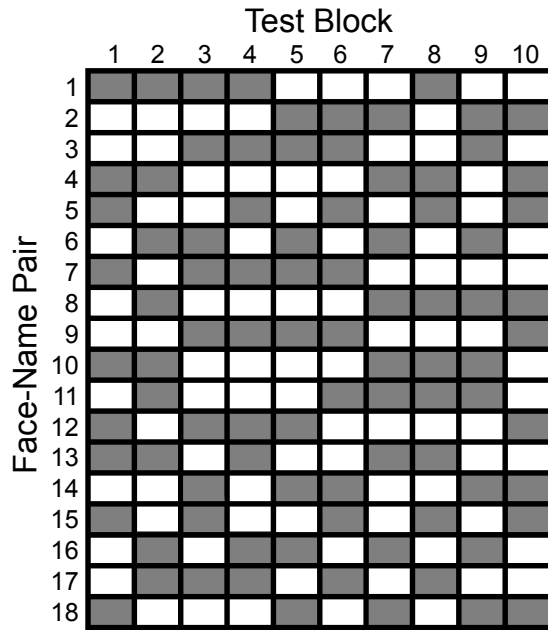
Figure 6. The status of the 18 study face-name pairs over the ten Test blocks. A filled rectangle signifies that the studied face-name pair was preserved in that Test block (a *target* pair); an unfilled cell signifies that the study face-name pair was rearranged in that Test block (a *lure* pair). Each study pair was preserved in five of the ten Test blocks, with each Test block comprising nine *target* pairs and nine *lure* pairs.
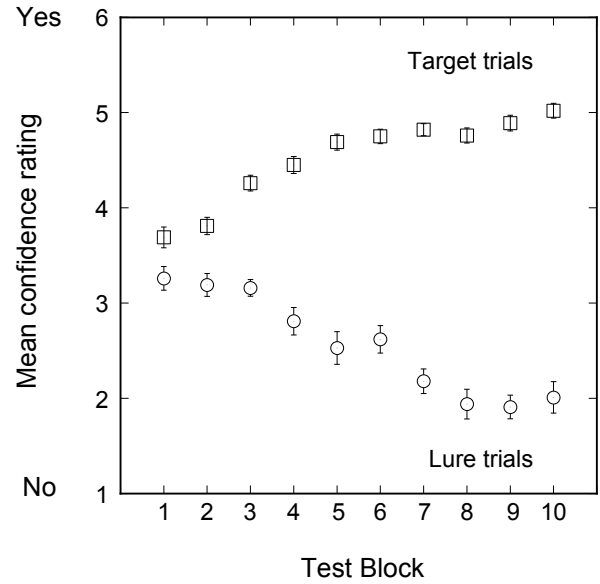

Figure 7. Mean confidence ratings elicited by *Target* (□) and *Lure* (○) stimuli over 10 successive Test blocks. Higher ratings represent greater confidence that the stimulus pair was one that had been studied. Error bars are ±1 within-subject standard error of the means.

Results and Discussion

Subjects' performance in the same-different task confirmed that they readily distinguished among the synthetic faces, and among the spoken Chinese names. Across the 30 subjects, the"same"-"different" judgments for faces were correct on 92.3% (SD =1.4%) of the trials; the figure for names was 97.6% (SD = 2.2%). It might be surprising that the subjects, non-Asian speakers, managed to discriminate among the Chinese names so well. Recall, however, that we deliberately chose names whose phonological differences (on the second tone) would be most distinctive to Western, non-Chinese speakers (Shen, 1989). Subjects high levels of discrimination in judging items within each class of stimuli makes it unlikely that within-set confusions among faces or among names would be the primary cause of failures of associative recognition task.

As a result of the rating scale's bipolarity, high numerical ratings to a *Target* pair and low numerical ratings to a *Lure* pair were both correct responses. For some analyses below, we express results in terms of the numerical ratings. Other, model-related analyses were better served by less detailed equivalents to "Yes"–"No" recognition judgements. For such analyses, ratings of 4, 5, or 6 were taken as "*Yes*"; ratings of 1, 2, or 3 were taken as "*No*" responses.

To gauge improvement in associative recognition with repeated opportunities to study the same face-name pairs, we examined subjects' rating responses over successive test blocks. Figure 7 shows the mean ratings for *Target* trials (■) and *Lure* trials (○) as a function of test block. As expected, average recognition accuracy increased substantially over test blocks: ratings for target pairs also increased significantly, while ratings for lure pairs decreased significantly ($F(9,261) = 26.61, p < 0.001$, and $F(9,261) = 28.13, p < 0.001$, respectively). These changes correspond to a change in recognition accuracy ("yes"-"no" recognition judgments) from 0.55 (SD = 0.13) in Test Block 1, to 0.82 (SD = 0.14) in Test Block 10 (one-way ANOVA, $F(9,261) = 33.70, p < 0.001$). Both measures of change agree that repeated presentations strengthened associative recognition for face-name pairs, that is, ratings for *Target* pairs grow while ratings for *Lure* pairs decline.

Effects of summed-similarity

We analyzed the effects of summed-similarity separately for *Target* pairs and for *Lure* pairs. The left panel in Figure 8 shows the mean numerical ratings for target pairs as a function of face summed-similarity. The parameter for the family of curves is name summed-similarity. The effects of face summed-similarity and the effect of name summed-similarity were each statistically significant (two-way repeated measures ANOVA; see Table 3). Subjects generally gave lower ratings to *Target* pairs whose face's summed-similarity was high, and to *Target* pairs whose name's summed-similarity was high ($F(2,116) = 4.90, p = 0.011$ and $F(2,119) = 23.48, p < 0.001$, respectively). That

Table 3
Results of two-way repeated measures ANOVAs analyzing the effects of summed-similarity of faces (low, median and high) and summed-similarity of names (low, median and high) on three dependent variables: the confidence judgment for target pairs, that for lure pairs, and Ag – the area beneath ROC curve.

Dependent Variable	Face Summed Similarity		Name Summed Similarity		Face × Name	
	$F(2,116)$	MSe	$F(2,116)$	MSe	$F(4,116)$	MSe
Confidence, target pairs	4.90^*	1.50	23.48^{***}	7.87	8.66^{***}	2.90
Confidence, lure pairs	9.10^{***}	1.58	29.85^{***}	4.61	1.13^{ns}	0.12
Ag	20.82^{***}	0.23	3.64^*	0.039	3.37^*	0.035

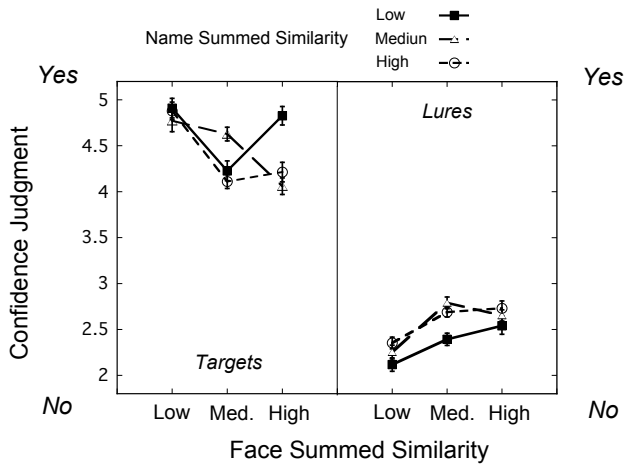$^{ns}\ p > 0.1.$ $^*p < 0.05.$ $^{***}p < 0.001.$

Figure 8. Mean confidence ratings vs. the summed-similarity of face stimuli. Values were aggregated across all test blocks and subjects. Separate curves represent different levels (high, medium and low) for name stimuli summed similarity. The results of target pairs were plotted on the left, and the results of lure pairs were on the right. Error bars are ±1 within-subject standard error of the means.

is, subjects were less confident in recognizing a *Target* pair when its constituent items were similar to other items in the respecitve stimulus classes. The inverse pattern was seen for numerical ratings of *Lure* pairs. As the right-hand panel of Figure 8 shows, increasing face summed-similarity or increasing name summed-similarity increases the numerical ratings given to *Lure* pairs ($F(2,116) = 9.10, p < 0.001$ and $F(2,116) = 29.85, p < 0.001$, respectively). In other words, when either item in a *Lure* was similar to studied items of the same class, subjects were less certain about rejecting the false, re-arranged pair.

For a converging related perspective on these results, we examined the impact of face and name similarity via *Ag*, a measure of recognition that should be less susceptible to variation in the criterion with which information in memory was translated into recognition responses. This index of area under the ROC confirmed what was noted in the recognition responses alone. *Ag* showed a significant

decrease as summed-similarity increased, and did so for both face summed-similarity and for name summed-similarity ($F(2,116) = 20.82, p < 0.001$ ($M_{low}= 0.84$, $M_{medium} = 0.75$, $M_{high} = 0.76$), and $F(2,116) = 3.64, p = 0.032$, ($M_{low} = 0.80$, $M_{medium} = 0.78$, $M_{high} = 0.76$), respectively. Together, the results suggest that face similarity and name similarity each had a distinct influence on associative recognition, a face-name pair with a higher summed-similarity stimulus (face or name) yielded *poorer* recognition performance.

Ratings reflect summed-similarity

To assess the dependence of subjects' ratings upon summed-similarity for faces, we regressed those ratings against the summed Euclidean distance in MDS space between a pair's face stimulus and all the other study faces. An equivalent analysis was done for the name stimulus in each pair. As explained earlier, we did not have a direct measure of a stimulus' summed-similarity to the other stimuli in its class. As a result, in the following regression analysis, values of summed-pairwise distances serve as surrogates for values of summed-similarity. As mentioned above, we assume that summed-similarity would be an inverse, monotonic function of summed-distance.

Target pairs The regression slope coefficients (β values) showed a *positive* relationship between rating judgments and a face's summed similarity relative to all 18 studied faces. As shown by the solid markers in Figure 9A, the higher a face's summed-distance, the more strongly did subjects endorse a studied test pair, which included that face stimulus. In particular, 95% confidence intervals around the β values was [0.043, 0.12], significantly above zero, $t(29) = 4.47, p < 0.001$. The same pattern was seen for the name stimuli in *Target* pairs. The empty marker in Figure 9A shows that β values for name stimuli were significantly above zero; the 95% confidence interval was [0.042, 0.10], $t(29) = 4.86, p < 0.001$. Given the inverse relationship between summed-distance and summed-similarity, these regression analyses, for faces and for names, are consistent with the result seen in the lefthand panel of Figure 8, namely that a *Target* face-name pair with a high summed-similarity component, either face or name, tends to be less confidently recognized.
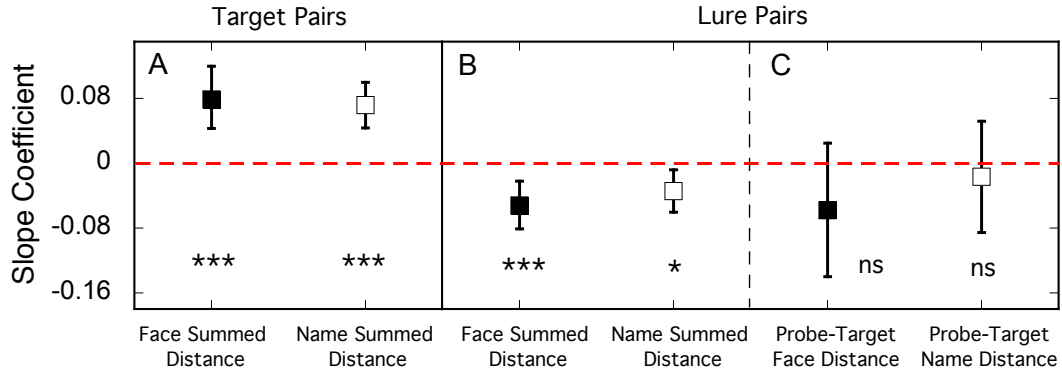
Figure 9. Mean slope coefficients (β values) from several regressions. The horizontal dashed line corresponds to β=0. A: Rating judgments for target pairs were regressed separately on face summed-distance (solid marker), and on name summed-distance (empty marker). B: Rating judgments for lure pairs were regressed separately on face summed-distance (solid marker), and on name summed-distance (empty marker). C: β values produced when rating judgments for lure pairs were regressed against the probe-target distance of the pair's face stimulus (solid marker), and name stimulus (empty marker). Error bars are 95% confidence intervals. Results of one-sample *t*-tests: $*** = p < 0.001$, $* = p < 0.05$, and *ns* = non-significant.

Lure pairs As predicted, β values from regressions for *Lure* stimuli exhibited a relationship between subjects' judgments and an item's summed-distance that differed from what was seen for *Target* pairs. In particular, for both faces and names ratings tended to be inversely related to an item's summed-distance from items of the same class in the studied set. Figure 9B shows that for both names and faces, β values tended to be less than zero: for face summed-distance, the 95% confidence interval for the slopes was $[-0.081, -0.022]$, $t(29) = -3.60, p < 0.001$; for name summed-distance, the 95% confidence interval was $[-0.060, -0.008]$, $t(29) = -2.68, p = 0.012$. Taking into account the inverse relationship between summed-distance and summed-similarity, Figure 9B's result is consistent with the finding shown before, in the righthand panel of Figure 8: The relationship between numerical ratings and summed-distance values for lure pairs suggests that subjects often falsely endorse a lure pair, and do so with remarkable confidence, if the pair's face or name item is highly-similar to other studied items in its class.

Probe-target distance alone does not control recognition

Our analyses of similarity's effect on face-name learning and global matching accounts of recognition share a key assumption, namely that summed-similarity (or its inverse, summed-distance) is the appropriate metric for similarity (Clark & Gronlund, 1996; Sekuler & Kahana, 2007). Before proceeding to examine model-based accounts of our results, it seemed prudent to compare this similarity metric against a plausible alternative. So, we asked whether the correlation for *Lure* pairs between summed-distance and ratings (see Figure 9B) might reflect something other than the summed-distance between the probe pair and all the other studied pairs. We wondered if, instead, this correlation reflected the (i) distance between each item in a *Lure* pair

and the distance between it and the item with which it had been studied. For each item in the *Lure* pair, whether it be a name or a face, we call this distance the *probe-target distance*. The idea that the single probe-target distance would be consequential has intuitive appeal. After all, when ratings for lure pairs were regressed against summed-distance, any value of the <u>summed</u>-distance necessarily had to include the *probe-target distance*. Thus, it might be that what the preceding paragraph attributed to summed-distance actually arose solely from the distance between probe components and mates with which they had been studied. Consistent with that possibility, Pantelis et al. (2008) found that "*Yes*" responses to rearranged face-name pairs (*lures*) increased if the probe face looked similar to a different target face that had been previously paired with the probe name.

To examine the effect of probe-target face distance, we regressed subjects' ratings on *lure* trials against the probe-target distance of faces in lure pairs. In addition, we carried out analogous regressions with name probe-target distance in lure pairs.[3]

For each of the *Lure* pairs presented in the 10 Test blocks, we computed probe-target distances for that pair's name and for the pair's face. For example, consider a *Lure* pair (f_j, n_k) whose component face, f_j, during Study blocks was paired with name n_j, while its component name n_k was paired with face f_k. For the face in *Lure* pair (f_j, n_k), the face probe-target distance is given by the Euclidean distance in MDS space between f_j and f_k. Similarly, for the name in that same *Lure* pair, the name's probe-target distance is given by the distance between n_k and n_j. For individual subjects, the ratings given to various *Lure* pairs was regressed on the distance between probe and *Target* for faces and for names.

---

[3] As target pairs entailed correct, preserved pairings, the probe face in a target pair was the target face, as were the probe and target names. Therefore, the regression on probe-target distance was not done for target pairs.

Then, for faces and for names, we assessed the β values produced by the regressions.

There was no reliable relation between subjects' ratings and the distance in MDS space that separated probe from target. As shown by the solid and empty markers in Figure 9C, β values for both names and faces were not significantly different from zero. Specifically, the 95% confidence intervals from the regressions for probe-target distances were $[-0.14, 0.025]$ and $[-0.086, 0.052]$, for faces and names, respectively ($p = 0.15$ and $p = 0.50$). These non-significant regressions contradict the idea probe-target distance, rather than summed-pairwise distances is actually responsible for the significant relationships shown in Figure 9B.

Model-based analyses of associative recognition

We performed two distinct model-based analyses in order to identify the processes responsible for associative recognition. We began with the assumption that any model should satisfy two key constraints from our empirical findings: first, that recognition of a probe pair reflects the pair's summed similarity to studied pairs; second, that recognition's dependence upon summed similarity should operate in opposite directions for *Target* and *Lure* probes. This second constraint is based the pattern of results in Figure 8.

To gauge its ability to account for our results, we fit NEMo, a summed-similarity model of recognition, to the empirical values of p(Yes) as derived from subjects' confidence ratings, as explained earlier. To focus on the mechanisms of the observed similarity effects, we averaged subjects' responses over all 10 Test blocks. This effectively sets aside any examination of changes in performance over successive Test blocks, which would be an interesting topic for future research, but is not directly relevant to our concern here.

For *Target* pairs, we found the mean p(Yes) value for each of the 18 study pairs, two exemplars within each of the nine combinations of face and name summed similarity. Then we averaged the values associated with each exemplar pair. This produced a mean p(Yes) value for each combination of face-name similarity (*e.g.*, high face summed similarity and medium name summed similarity). For *Lure* pairs, a slightly different approach was required. As each unique *Lure* pair appeared only once during testing, we averaged subjects' p(Yes) responses for each lure pair, then as we did with *Target* pairs, p(Yes) values were averaged for *Lures* each similarity category (3 levels of face summed similarity × 3 levels of name summed similarity).

NEMo bases its recognition responses on the summed similarity between a test probe, p, and noisy exemplars in memory of each study item, $s_i + \varepsilon_i$ (with ε_i representing the noise parameter). When NEMo's summed similarity exceeds a decision criterion (C), NEMo judges p to have been in the set of study items; otherwise the model declares p a new, non-studied item.

$$(2) \qquad \sum_{i=1}^{18} \eta(p, s_i + \varepsilon_i) > C$$

To compute the similarity values required to fit NEMo, we applied Equation 3 to the inter-item distances in the MDS solutions for faces and for names.

(3)
$$\eta(p, s_i + \varepsilon_i) = e^{-\tau d(p, s_i + \varepsilon_i)}$$
$$= e^{-\tau \sqrt{d_{face}(p, s_i + \varepsilon_i)^2 + d_{name}(p, s_i + \varepsilon_i)^2}}$$

For each category of summed similarity, NEMo was fit to mean p(Yes) values for *Target* and *Lure* probe pairs. The resulting 18 values, nine for *Targets* and nine for *Lures*, were fit by means of a multi-start, stochastic hill-climbing algorithm that optimized $ln(likelihood)$. The parameters that were estimated, along with the range of values explored for each, were τ (the exponential gradient of the similarity function as in Equation 2 and 3, [0, 3]), ε_f, ε_n (noise associated with the memory representations of faces, and of names, [0, 3]), and C (the decision criterion, [0, 2]). With optimized parameters. NEMo produced a respectable overall fit: $r^2 = 0.92$, and $RMSD = 0.063$ (see the scatter plot in Figure 10. However, as anticipated, NEMo predicted incorrectly that p(Yes) would increase with summed similarity for both categories of test stimuli, *Lures* (rearranged study items) and *Targets* (intact study pairs). As a result, NEMo failed to satisfy the second empirical constraint that we described earlier, a shortcoming whose detection was made possible by the systematic variation of stimulus similarity.

An alternate model for associative recognition: NEMo.D. Earlier, we cited two empirical constraints that a model had to satisfy. Although NEMo satisfied the first constraint, it failed to satisfy the second. Therefore, we expanded NEMo to incorporate a recollective process that functions in tandem with a global matching computation of probe familiarity. In this expanded model, NEMo.D, upon the presentation of a test probe, associative recognition either proceeds along a global matching (NEMo) route, as described earlier, or proceeds along what we can call the "Recollection Route" to recognition. Lest the term be misinterpreted, "recollection" is defined as the act of remembering something, but, of course, remembering does <u>not</u> guarantee the validity of what is remembered.

NEMo.D begins with a variant of Luce's choice rule (1959/2005). This choice rule formulation allows stimulus similarity to determine by which of the two alternative routes a test probe will be processed and judged. Given some test probe, p, the probability that p will be processed via the Recollection Route and result in the recollection of a study item s_k, depends upon p's similarity to s_k, and upon the
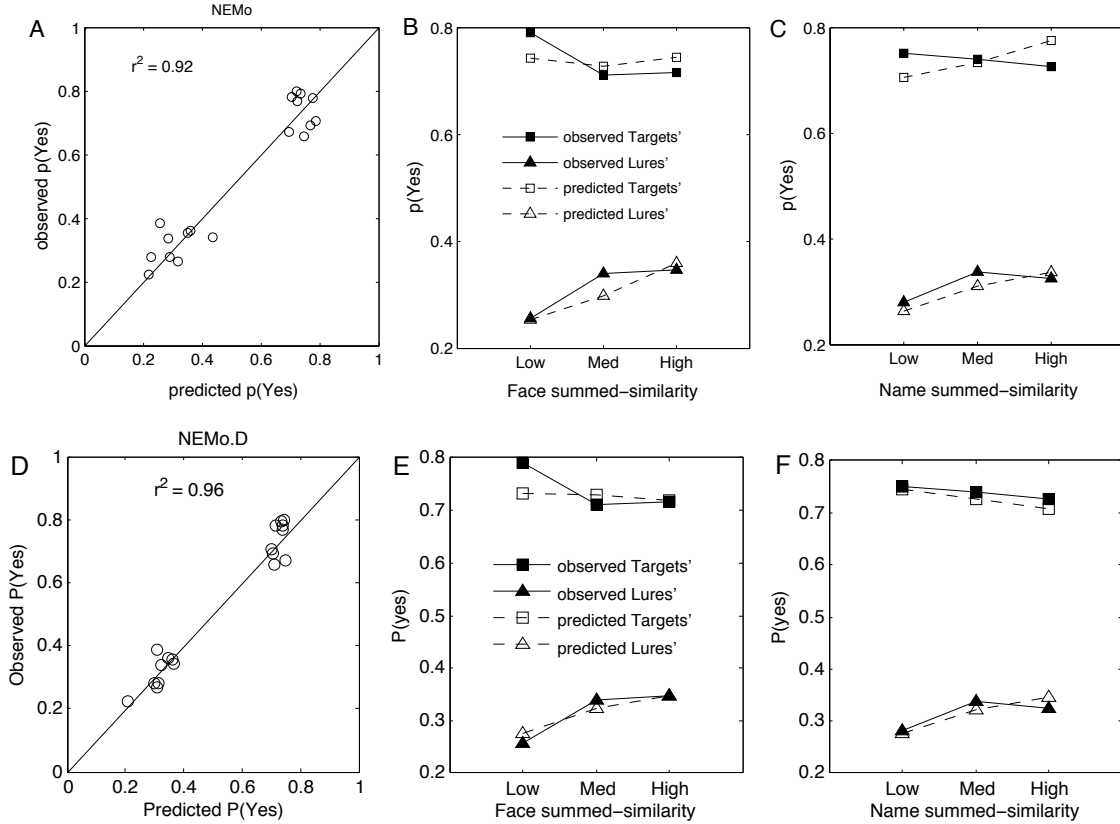
Figure 10. Panels A-C: Results of the best-fit NEMo model; Panels D-F: results from NEMo.D with optimized parameter values. Panel A: The predicted p(Yes) for recognition responses in the best-fit NEMo model are plotted against the observed p(Yes), $r^2 = 0.92$. Panel B: Observed (solid markers) and predicted (empty markers) p(Yes) for *Target* and *Lure* pairs are plotted as a function of *face* summed-similarity of a probe pair. Panel C: Observed (solid markers) and predicted (empty markers) p(Yes) for *Target* and *Lure* pairs as a function of *name* summed-similarity of a probe pair. Panel D: The predicted p(Yes) for recognition responses in the best-fit NEMo.D model are plotted against the observed p(Yes), $r^2 = 0.96$. Panel E: Observed (solid markers) and predicted (empty markers) p(Yes) for *Target* and *Lure* pairs are plotted as a function of *face* summed-similarity of a probe pair. Panel F: Observed (solid markers) and predicted (empty markers) p(Yes) for *Target* and *Lure* pairs as a function of *name* summed-similarity of a probe pair.

sum of p's pairwise similarities to each study item. More formally,

$$(4) \qquad pr(Recollection \to s_k) = \frac{\eta(s_k, p)}{\sum_{i=1}^{18} \eta(s_i, p) + h}$$

Note that the h parameter modulates the probability that test probe p is processed via one route rather than the other. The probability that test probe p will be processed via the NEMo route is given by

$$(5) \qquad pr(NEMo) = \frac{h}{\sum_{i=1}^{18} \eta(s_i, p) + h}$$

If a probe is processed via NEMo, the recognition response is controlled by processes summarized in Equation

2. If, however, a probe is processed via the alternative, Recollection route, Equation 4 implies that the probability that any particular studied pair, such as s_k, will be recollected, depends on its similarity to the probe, $\eta(p, s_k)$, and on the probe's summed similarity to all study items.

From Equation 4, we can see that if the test probe were p, and if $p \cong s_k$, the probability of recollecting that correct, p-congruent studied pair, s_k, would be

$$(6) \qquad \begin{aligned} pr(recollection \to s_k) &= \frac{\eta(s_k, p)}{\sum_{i=1}^{18} \eta(s_i, p) + h} \\ &= \frac{1}{\sum_{i=1}^{18} \eta(s_i, s_k) + h} \end{aligned}$$

Once some studied pair, say s_k, has been recollected, if that study pair matches p, the response is "*Yes*",

otherwise, the response is "*No*". As a result, if a *Lure* pair happens to be processed via the Recollection route, that *Lure* will be rejected as a match, producing a "*No*" response. As for *Target* pairs, only when the correct studied pair is recollected, does the model accept the *Target* pair as correctly paired and responds "*Yes*". Therefore, the probability of correctly endorsing a *Target* pair in recollection route is inversely related to the summed similarity of this *Target* pair to all the remaining study pairs (see Equation 6). More specifically, if a *Target* pair is highly similar to other study pairs, recollection will produce a relatively low p(Yes) to this target. Intuitively, this pattern would mimic the similarity effect on *Target* pairs that we saw in our subjects' recognition data. More importantly, this is precisely where NEMo, operating alone, failed.

Note that NEMo.D's choice rule operates adaptively, differentially affecting the likelihood on *Target* and on *Lure* trials that a probe will be processed via NEMo. According to Equation 5, the probability that a probe will be processed via NEMo is inversely related to the summed similarity of that test probe to all studied pairs. As a result, *Lures* are more likely to be processed via NEMo's global matching process (which predicts *higher* p(Yes) for pairs with higher summed similarity, consistent with *Lures*' result), while *Targets* are more likely to be processed via the Recollection route (which predicts *lower* p(Yes) for pairs with higher summed similarity, consistent with *Targets*' result).

We fit NEMo.D to subjects' 18 mean p(Yes) responses, averaged over the 10 Test blocks. The model was fit using a multi-start, stochastic hill-climbing algorithm that optimized $ln(likelihood)$. In addition to NEMo's four parameters, the NEMo.D includes the h parameter (see Equation 4). Values of h from 0 to 2 were searched, and the best fit was produced with $h = 0.57$.

The r^2 from the optimized parameter values for NEMo.D is 0.96, *RMSD* is 0.042, a reduction of ~50% from the value produced by NEMo alone. It should be noted that because our results cluster into two distinct groups, r^2 provides quite an insensitive assay of goodness of fit. Based on Equation 5 and the parameter values estimated for the best-fit model, over the 10 Test blocks, various *Lure* probe pairs went through NEMo 35.3–57.3% of the time on average; while *Target* pairs took NEMo less frequently, 25.0–28.4%.

Table 4
Values of parameters for best-fitting models

	NEMo	NEMo.D
τ	2.258	1.065
ε_{Faces}	0.427	0.535
ε_{Names}	1.063	0.838
Criterion	0.113	0.677
h	–	0.571

As mentioned earlier, a critical success of the model should be the ability to simulate similarity effect on recognition responses for targets and for lures. As can be seen in Figure 10's lower row of graphs, the predicted p(Yes)

of the best-fit NEMo.D model closely matches the observed data: the predicted p(Yes) for targets decreases with summed similarity, while the predicted p(Yes) for *Lures* increases with summed similarity.

To compare NEMo.D with a model in which NEMo alone determines performance, we calculated the Bayesian Information Criterion (BIC). This measure takes account of differences in the number of parameters associated with the competing models (Schwarz, 1978). Under the assumption that model errors are normally distributed,

$$(7) \qquad BIC = n \cdot ln\left(\frac{RSS}{n}\right) + k \cdot ln(n)$$

where k is the number of free parameters to be estimated, n is the number of observations, and RSS is the residual sum of squares from the estimated model. A better-fitting model produces lower BIC values. Even after imposing a penalty for the additional parameter NEMo.D significantly improves the goodness-of-fit significantly relative to NEMo ($BIC = -99.50$ and $BIC = -87.74$, respectively.)

As NEMo is nested within NEMo.D, we calculated $-2ln\lambda$ with NEMo.D and NEMo's *LLi*. This statistic is asymptotically distributed as χ^2, with degrees of freedom equal to the difference in the models' number of free parameters (df =1). A χ^2 test confirms that NEMo.D is significantly better than NEMo ($LLi = -76.63$ and $LLi = -105.05$, respectively, $p < 0.001$).

Finally, Figure 11 is meant to provide a clearer intuition of how the h parameter affects recognition. To generate the results shown in the figure, NEMo.D was re-run twice, using distinctly non-optimal values of the h parameter, one higher than the optimal value, one lower. While h varied, other parameter values were clamped at their respective values from the best-fit model. As Equation 4 suggests, the higher value of h caused more stimuli to be processed through NEMo; the lower value of h caused fewer stimuli to be processed through NEMo. As mentioned earlier, with $h = 0.57$ (the estimated optimal value), p(NEMo) for *Lure* stimuli ranged from 35–57%, and *Target* stimuli ranged from 25–28%. In contrast, when $h = 0.17$ (the value on which Figure 11A is based), stimuli were far less likely to processed via NEMo, namely 9–29 for *Lures* and 9–11% for *Targets*. With $h = 0.97$ (the value on which Figure 11B is based), stimuli were more likely to be processed via NEMo, 48– 69% and 36–40%, for *Lures* and *Targets*, respectively. Comparing the corresponding panels in Figure 11A and B confirms that, as expected, the probability of processing a probe pair via NEMo has a stronger effect on responses to *Lure* pairs than on responses to *Target* pairs. This differential effect reflects the fact that with the optimal value of h, *Lure* pairs are the more likely to be processed via NEMo than *Target* pairs. Finally, comparing the middle and rightmost panels shows that variation in h has very nearly equal modulating effects on the similarities of names and of faces. This result does not mean, of course, that face and name similarity will always operate in equal measure. Instead, these two sources of

similarity probably had quite comparable effects because the comparable similarity relationships among our face stimuli, on one hand, and among our name stimuli, on the other (see Figures S1 and S2).

General Discussion

We found that subjects' memory for face-name pairs depends systematically on the number of similar competitors within the set of studied names and faces. When an intact *Target* pair included either a face or a name that was similar to other studied stimuli, this pair was less confidently recognized as correctly paired. In contrast, when a rearranged *Lure* pair included a face or a name similar to other stimuli, it was more likely to be erroneously endorsed as having been studied before. The parallel results we observed for variation of visual similarity (for faces) and phonological similarity (for names) suggests that these findings reflect general similarity-based mechanisms underlying recognition memory.

The difference in the pattern of results observed for target and lures (see Figure 8) runs contrary to expectations of summed similarity models of associative recognition. If associative recognition were driven by the summed similarity of the cue-target pair to the pairs stored in memory the curves in both panels would increase with summed similarity. Whereas that is surely the case for *Lure* trials, the opposite pattern is observed on the *Target* trials.

This failure of summed-similarity theory suggests that associative recognition of face-name pairs is not process pure, and that something other than, or in addition to, global matching is at work. To test this idea formally, we implemented a dual-process model in which recognition decisions are based on either a familiarity or a recollection signal. We then fit this model to our experimental data. In our NEMo.D model, a test probe can evoke recollection of a specific previously studied pair and use the match between the probe and retrieved information to make a recognition judgment. Alternatively, if the similarity between the test probe and the studied item is not sufficient to retrieve a specific prior pair, the model basis its recognition decision on the familiarity of the test probe, which is operationalized as the summed-similarity between the probe and the contents of memory, as specified in the noisy exemplar model (NEMo) of perceptual item recognition (Kahana & Sekuler, 2002). NEMo.D assumes that the summed similarity between the test probe and the set of previously studied pairs determines the likelihood that one of the pairs will be recollected. If the total similarity is low, the model is more likely to base its decision on summed similarity whereas if the total similarity is high, the model is more likely to base its decision on the recollection of a specific pair.

Without a recollection component, our summed similarity model, NEMo, could not account for the similarity effects observed in our associative recognition experiment. In contrast, the dual-process implementation, NEMo.D, nicely captured all of the trends we observed. Unlike most prior dual process models of recognition memory, which are

bivariate generalizations of signal detection theory, often with a high threshold or infinite signal on the recollection dimension, NEMo.D assumes that both recollection and familiarity processes depend on the similarity relations between the probe item, or pair, and the memorial representations of the studied items. As such, our model can make specific predictions about how similarity should affect associative recognition at the level of individual probe items or pairs. In the case of familiarity, our model specifies where the variability in memory "strength" comes: it arises from the variable similarity relations between the probe and the contents of memory. In the case of recollection, our model specifies how subjects may "falsely" recollect an item or pair if the item or pair are similar to the probe item. The similarity-based familiarity and recollection mechanisms in NEMo.D also provide a nice bridge between dual process signal detection models of associative recognition and more fully specified attribute-based models of recall and recognition (e.g., Nosofsky et al., 2011; Dennis & Humphreys, 2001; Howard & Kahana, 2002; Shiffrin & Steyvers, 1997).

The summed-similarity mechanism used to model the familiarity signal in traditional dual-process models may also help to explain distinctiveness effects in name-face learning. Distinctiveness can be operationalized in various ways. For example, the distinctiveness of faces can be operationalized as ones that subjects rate as easy to spot in a crowd (Valentine, 1991; Valentine & Bruce, 1986; Hancock, Burton, & Bruce, 1996; Hosie & Milne, 1996). Valentine (1991) suggested a distinctive face is one that is relatively isolated in faces' multidimensional similarity space. In quantitative terms, Loffler et al. (2005) defined a face's distinctiveness by its distance from the mean face in face multidimensional space. It is important to recognize that although distinctive faces will tend to have lower summed similarity than more prototypical faces, the two measures are quite different. The difference arises from the non-linearity in the function relating similarity to distance (Equation 1). This non-linearity implies that a face that is very close to a small number of faces but far from the mean could have very high similarity, whereas a face that was exactly equal to the mean, but not close to any individual faces would have low summed similarity.

The idea that summed similarity underlies familiarity is also supported by findings that the shape of the receiver-operating characteristic (ROC) curve for visual textures is influenced by summed similarity (Yotsumoto, Kahana, McLaughlin, & Sekuler, 2008). Because NEMo.D assumes that both familiarity and recollection are determined by the similarity relations among the probe and the contents of memory, the model offers a clear explanation for the variability in performance across different items (Gold, Murray, Sekuler, Bennett, & Sekuler, 2005; Huang & Sekuler, 2010). Although we have not attempted to measure directly the noise associated with individual stimuli, techniques for such measurements are available and could be exploited in future work (Pelli & Farell, 1999; Ahumada, 2002).
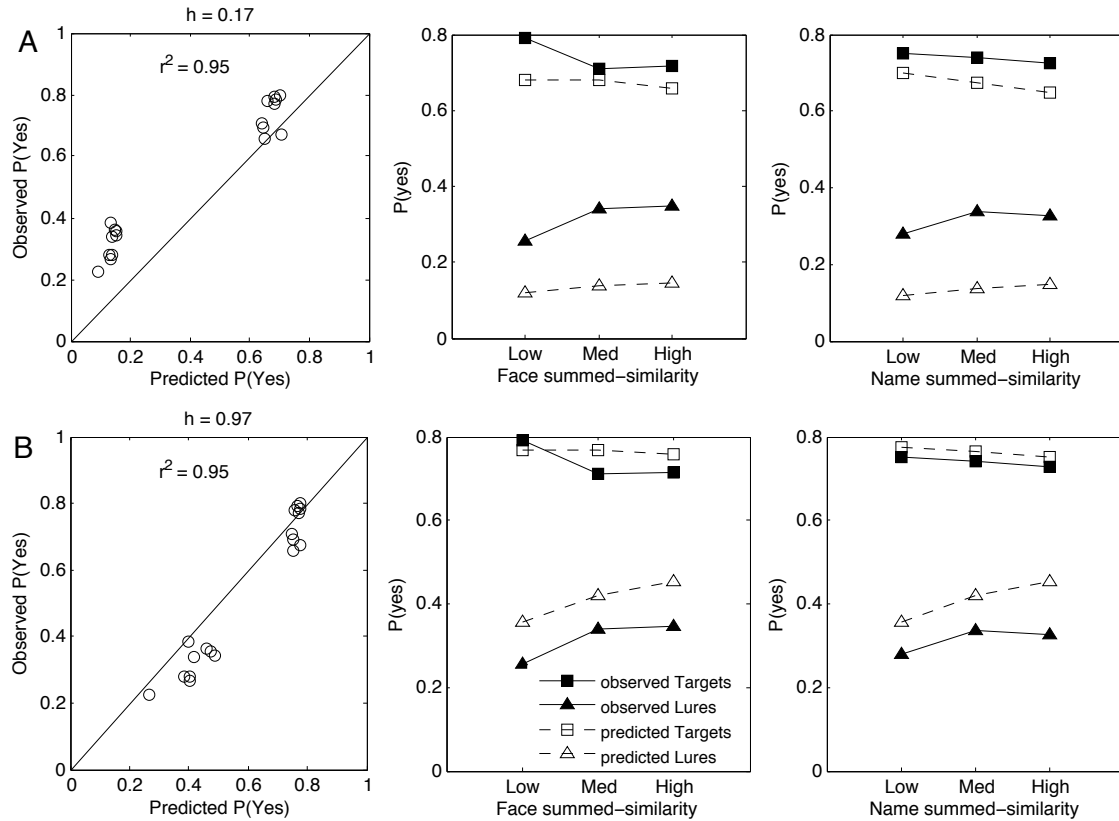
Figure 11. Fits to data with non-optimized values of the *h* parameter in NEMo.D. Panel A: *h*=0.17; Panel B: *h*=0.97. For both cases, parameters other than *h* were fixed to their respective values from the best-fitting model. Note that for both name and face components of a pair, variation in *h* changes predictions for *Lure* pairs (△) far more than it does for *Target* pairs (□).

Lessons and limitations

An important future direction will be to extend NEMo.D to multi-trial learning effects such as the ones represented in Figure 7. Unfortunately, Experiment Two's design would present inherent obstacles to such an effort. The first obstacle lay in the fact that all subjects shared the same sequence of faces and names during successive study blocks (see Figure 5B); the second obstacle arose from the fact that successive test blocks differed in the specific items with which subjects were tested (see Figure 6). The first of these obstacles sharply limited our ability to generalize from this particular sequence of learning opportunities to statements about face-name learning with other sequences; the second of the obstacles made it difficult to compare performance on successive test blocks –as the test items differed from one block of trials to another. Clearly, any study whose primarily focus was on learning must avoid both these obstacles. Many previous studies have examined face-name learning, but few have examined the effects of similarity, and fewer still have derived their measurements of learning from an associative recognition task.

Given the extensive psychological evidence supporting the distinction between item and associative information

(Murdock, 1974; Malmberg & Xu, 2007; Kahana, 2012) it is not too surprising that modeling similarity effects in associative recognition required that both summed-similarity and associative recognition mechanisms. Indeed, neural evidence from both human and animals also strongly supports this distinction (Diana, Yonelinas, & Ranganath, 2010; Davachi, 2006; Song, Jeneson, & Squire, 2011; Squire, Wixted, & Clark, 2007). By exploiting as stimuli names and faces whose similarity-relations were well quantified, our experimental analysis of recognition memory for name-face associations revealed a pattern that violated predictions of well-established summed similarity models. We show that these similarity effects are neatly accounted for by NEMo.D, a dual-process model of associative recognition that combines a summed-similarity familiarity process with a probabilistic similarity-based recall process. Although our first step was to apply NEMo.D to memory for name-face associations, our model can be easily applied to any stimulus domain for which the similarities among items can be quantified, including simple visual forms, complex scenes, and even words. In all of these cases, NEMo.D will make predictions about the variation in performance across individual probe items.

References

Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*(1), 121–131.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and application* (2nd ed.). New York: Springer.

Bower, G. H., Thompson-Schill, S., & Tulving, E. (1994). Reducing retroactive interference: An interference analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 51–66.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.

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.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693–700.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–477.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience*, *22*(8), 1808–1818.

Fisher, R. A., & Yates, F. (1963). *Statistical tables for biological, agricultural, and medical research (6th ed., revised and enlarged).* New York: Hafner Publishiong Co.

Fraas, M., Lockwood, J., Neils-Strunjas, J., Shidler, M., Krikorian, R., & Weiler, E. (2002). 'What's his name?' A comparison of elderly participants' and undergraduate students' misnamings. *Archives of Gerontology & Geriatrics*, *34*, 155–165.

Galster, M., Kahana, M. J., Wilson, H. R., & Sekuler, R. (2009). Processing of facial identity and emotion: Perception and memory use similar structures. *Cognitive, Affective & Behavioral Neuroscience*, *9*, 412–426.

Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science*, *16*(10), 769–774.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory & Cognition*, *24*(1), 26–40.

Hosie, J., & Milne, A. B. (1996). The effect of experimental design on memory for typical and distinctive faces. *Memory*, *4*(2), 175–197.

Howard, M., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299.

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

Isaacowitz, D. M., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2006). Selective preference in visual fixation away from negative images in old age? an eye-tracking study. *Psychology & Aging*, *21*(1), 40–48.

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

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

Lea, M. A., Thomas, R. D., Lamkin, N. A., & Bell, A. (2007). Who do you look like? evidence of facial stereotypes for male names. *Psychonomic Bulletin & Review*, *14*(5), 901–907.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, *8*, 1386–1390.

Luce, R. D. (1959/2005). *Individual choice behavior: A theoretical analysis.* New York: Wiley (reprinted by Dover Publications).

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Malmberg, K. J. (2008). Recognition memory: a review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335–384.

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition*, *35*, 545–556.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum Associates.

Nelson, D. L., Bajo, M., McEvoy, C. L., & Schreiber, T. A. (1989). Prior knowledge: The effects of natural category size on memory for implicitly encoded concepts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 957–967.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315.

Pantelis, P. C., Vugt, M. K. van, Sekuler, R., Wilson, H. R., & Kahana, M. J. (2008). Why are some people's names easier to learn than others? the effects of face similarity on memory for face-name associations. *Memory & Cognition*, *36*(6), 1182–1195.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *16*, 647–653.

Roediger, r., H L, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*(803-814).

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, *16*(6), 305-310.

Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association*, *24*, 27–47.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychological Bulletin & Review*, *4*, 145–166.

Song, Z., Jeneson, A., & Squire, L. R. (2011). Medial temporal lobe function and recognition memory: a novel approach to separating the contribution of recollection and familiarity. *Journal of Neuroscience*, *31*, 16026–16032.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews Neuroscience*, *8*, 872–883.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: J Wiley & Sons.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161–204.

Valentine, T., & Bruce, V. (1986). Recognising familiar faces: The role of distinctivenss and familiarity. *Canadian Journal of Psychology*, *40*, 300–305.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Vol. 10). Newbury Park, CA: Sage Publications.

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, *42*, 2909–2923.

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

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

Supplementary Material

MDS-generated three-dimensional similarity spaces for faces and names.
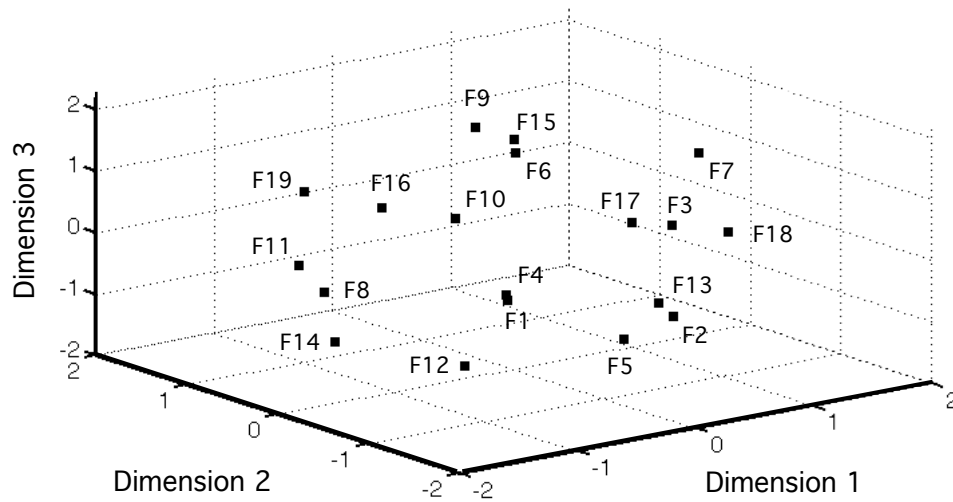

Figure 1. Experiment One's faces displayed in their three-dimensional similarity space. Dimension 1 was associated with head and chin shape; Dimension 2 with nose length and width; and Dimension 3 with mouth width and thickness of the upper lip.
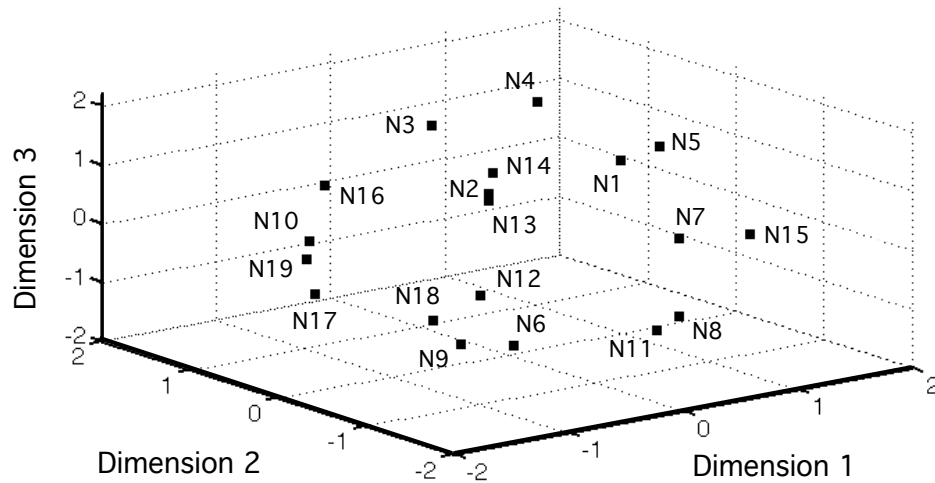

Figure 2. Experiment One's names plotted in their three-dimensional similarity space. Dimension 1 correlates with the phonological combination of the initial the component "i" in the final; Dimension 2 correlates with the phonological characteristics of the initial in the name; and Dimension 3 correlates with the phonological characteristics of the name's final, particularly the component "n".