Memory & Cognition
2007, 35 (6), 1233-1244

Recognition memory for realistic synthetic faces

Yuko Yotsumoto
Brandeis University, Waltham, Massachusetts

Michael J. Kahana
University of Pennsylvania, Philadelphia, Pennsylvania

Hugh R. Wilson
York University, Toronto, Ontario, Canada

and

Robert Sekuler
Brandeis University, Waltham, Massachusetts

A series of experiments examined short-term recognition memory for trios of briefly presented, synthetic human faces derived from three real human faces. The stimuli were a graded series of faces, which differed by varying known amounts from the face of the average female. Faces based on each of the three real faces were transformed so as to lie along orthogonal axes in a 3-D face space. Experiment 1 showed that the synthetic faces' perceptual similarity structure strongly influenced recognition memory. Results were fit by a noisy exemplar model (NEMO) of perceptual recognition memory (Kahana & Sekuler, 2002). The fits revealed that recognition memory was influenced both by the similarity of the probe to the series items and by the similarities among the series items themselves. Nonmetric multidimensional scaling (MDS) showed that the faces' perceptual representations largely preserved the 3-D space in which the face stimuli were arrayed. NEMO gave a better account of the results when similarity was defined as perceptual MDS similarity, rather than as the physical proximity of one face to another. Experiment 2 confirmed the importance of within-list homogeneity directly, without mediation of a model. We discuss the affinities and differences between visual memory for synthetic faces and memory for simpler stimuli.

Kahana and Sekuler (2002) developed a computational model that successfully accounts for short-term recognition memory with low-dimensional stimuli, compound sinusoidal gratings whose spatial frequency and phase vary. Building on Nosofsky's (1984, 1986) generalized context model (GCM), Kahana and Sekuler's noisy exemplar model (NEMO) combines core aspects of GCM with new key assumptions. NEMO follows the tradition of multidimensional signal detection theory (e.g., Ashby & Maddox, 1998) in assuming that stimulus representations are coded in a noisy manner, with different levels of noise associated with various dimensions. NEMO augments the summed-similarity framework of item recognition (Clark & Gronlund, 1996; Humphreys, Pike, Bain, & Tehan, 1989; Nosofsky, 1991, 1992) with the idea that recognition decisions are influenced not only by probe-to-list-item similarity, but also by the similarity of list items to one another, a variable that is called *within-list homogeneity*. Specifically, subjects appear to interpret probe-to-list similarity in light of within-list homogeneity, with greater homogeneity leading to a greater tendency to reject lures that are similar to one or more of the studied items. This impact of within-list homogeneity has been confirmed by Nosofsky and Kantner (2006), using color patches as stimuli, and by Kahana, Zhou, Geller, and Sekuler (2007), using compound gratings that were adjusted to reflect individual subjects' visual thresholds.

In contrast to compound sinusoidal gratings, essential aspects of visual processing of human faces take place several synapses beyond the primary visual cortex (Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005; Loffler, Yourganov, Wilkinson, & Wilson, 2005). Because the primary visual cortex participates not only in visual encoding, but also in visual memory and related phenomena (Klein, Paradis, Poline, Kosslyn, & Le Bihan, 2000; Kosslyn, Thompson, Kim, & Alpert, 1995; Magnussen & Greenlee, 1999), differences between visual processing of compound gratings and of human faces might produce corresponding differences in recognition memory for the two kinds of stimuli (Hole, 1996) and, thereby, undermine NEMO's applicability to face stimuli. Our aim here is not to explore or adjudicate among competing psychophysical or neural accounts of face perception and/or memory for faces (e.g., Gauthier, Skudlarski, Gore, & Anderson,

R. Sekuler, sekuler@brandeis.edu

2000; Grill-Spector, Knouf, & Kanwisher, 2004; Joseph & Gathers, 2002; Riesenhuber, Jarudi, Gilad, & Sinha, 2004) but to test NEMO's extensibility to memory for high-dimensional stimuli and to refine the methodology for measuring and modeling short-term memory.

Studies of face perception and/or face memory have used a range of stimuli, including simple cartoons, such as the Brunswik faces (Brunswik & Reiter, 1937; Peters, Gabbiani, & Koch, 2003; Sigala, Gabbiani, & Logothetis, 2002), photographs collected in convenience samples from sources such as school yearbooks, or images whose properties have been tailored to the study's specific purposes (e.g., Gold, Bennett, & Sekuler, 1999). We chose to work with realistic, synthetic human faces generated using methods introduced by Wilson, Loffler, and Wilkinson (2002). As test stimuli, Wilson faces mitigate problems arising from the availability of nonfacial information or from the availability of distinctive featural differences among faces (e.g., Duchaine & Weidenfeld, 2003; Sadr, Jarudi, & Sinha, 2003). Most important for model-related research, the generating algorithm for Wilson faces makes it easy to manipulate perceptual differences among faces. Furthermore, sets of Wilson faces can be represented in n-dimensional perceptual spaces whose properties can be tailored to suit some particular theoretical objective. For example, the axes of a face space might be orthogonalized, or individual exemplar faces might be made equidistant from some mean or reference face. Also, small, graded differences between faces make it difficult for subjects to learn and name each face in a reliable, consistent fashion. This is important because naming can subvert mnemonic reliance on visual information (e.g., Ashby & Ell, 2001; Goldstein & Chance, 1971; Hwang et al., 2005). To reinforce reliance on visual information per se, we limited rehearsal by permitting subjects only a brief glimpse of each face and then allowing only a short interval between successive faces. Wilson's synthesized faces are geometrically simple, as compared with actual, unprocessed grayscale photographs of faces, but these synthesized faces manage to convey sufficient information to permit identification of individual faces (Wilson et al., 2002). This individuality, which is important in episodic memory, is absent from some commonly used face stimuli, such as Brunswik faces (Brunswik & Reiter, 1937).

To preview, Experiment 1's design allowed us to apply the NEMO model of visual short-term recognition memory to performance on individual stimulus lists (i.e., diverse series of stimulus items). The model was applied in alternative modes—for example, expressing *similarity* either in terms of faces' physical coordinates or in terms of perceptual coordinates, assessed using multidimensional scaling (MDS). The design of Experiment 2 provided a direct, model-free demonstration of within-list homogeneity's influence on recognition decisions.

EXPERIMENT 1

Perceptual similarity among stimuli plays a central role in the structure of NEMO, and in other models of recognition as well. With compound gratings as stimuli, similarity has been defined by representing stimuli in a metric based on subjects' discrimination thresholds for spatial frequency (Kahana et al., 2007; Zhou, Kahana, & Sekuler, 2004). That same approach would likely fail with face stimuli because the representational space is clearly anisotropic. For example, the difference threshold for discriminating between faces varies substantially from one region of face space to another, with smallest difference thresholds in the neighborhood of the average face (Wilson et al., 2002). To permit full expression of potential anisotropies, we used nonmetric MDS to characterize the perceptual space within which our face stimuli were located and to quantify the distances between faces in that space. The data used for MDS came from oddity judgments made on simultaneously presented trios of faces.

Method

Subjects. Two male and 6 female volunteers from 20 to 25 years of age participated in the main experiment. All were naive as to the purpose of this experiment. They had normal or corrected-to-normal visual acuity, as measured with Snellen targets, and normal contrast sensitivity, as measured with Pelli–Robson charts (Pelli, Robson, & Wilkins, 1988).

Stimuli. Stimuli were generated and displayed using MATLAB and extensions from the Psychophysics and Video Toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 15-in. computer monitor with a refresh rate of 95 Hz and resolution set to 800 × 600 pixels. Routines from the Video Toolbox calibrated and linearized the display. Mean screen luminance was fixed at 36 cd/m².

The Wilson faces used in all our experiments were based on photographs of three Caucasian females whom we designate A, B, and C. From $\mathbf{m}_A$, the vector of 37 measurements taken on actual Face A, we synthesized a realistic version of that face in a stimulus space of high dimensionality ($n = 37$). Vectors of measurements taken on Faces A, B, and C were transformed mathematically so as to be mutually orthogonal, by Gram–Schmidt orthogonalization (Diamantaras & Kung, 1996; Principle, Euliano, & Lefebvre, 2000). Consequently, variation in one face's geometric properties is independent of the variation in geometric properties of the other two faces (Wilson et al., 2002). Additional details of the faces' construction and properties are given in the Appendix.

After preprocessing and normalization, vectors of measurements from several different faces can be combined to generate $\mathbf{m}_{avg}$, the vector of measurements for an *average* face. To illustrate, the synthesized mean female face is shown at the left of Figure 1. This mean face is derived from a sample of 40 Caucasian female faces. Summing $k(\mathbf{m}_{avg})$ and $(1-k)(\mathbf{m}_A)$, for some k, $0 < k < 1$, generates a face that is a mixture of the mean face and some particular individual face—in this case, A. Allowing k to vary, $k = 0 \ldots 1$, generates a graded series of faces spanning a continuum from the mean synthesized face (when $k = 1$) to a synthesized version of Face A alone (when $k = 0$). The same operation also can generate a graded series of faces, which span the distance from the mean face toward any other face—here, toward B or C. With these faces, it is easy to vary one stimulus' similarity to another, a variable that is important in recognition memory and central to NEMO.

The graded series for Faces A–C are shown in the upper three rows of Figure 1. Within each row, the value $(1-k)$ ranges from .04 to .20, in increments of .04. This means that each face differs from its nearest neighbor by approximately the mean discrimination threshold taken under viewing conditions similar to the ones we used (Wilson et al., 2002). In geometric terms, Faces A–C lie along the mutually perpendicular axes of a 3-D space, with the mean face at the origin.

A final set of faces, *D*, was generated by averaging corresponding exemplars of A, B, and C. The resulting faces are shown in the bottom row of Figure 1. Geometrically, the faces in row D lie along the diagonal

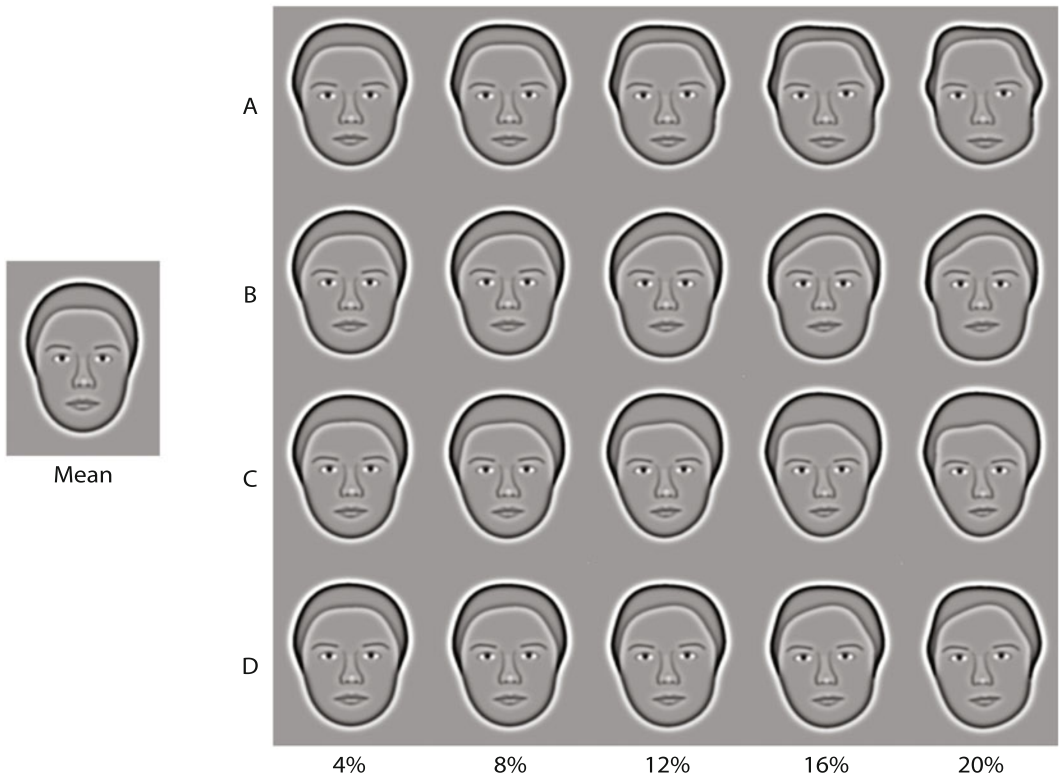

Figure 1. Face stimuli used in Experiment 1. Constructed after the method of Wilson, Loffler, and Wilkinson (2002), the stimulus labeled *mean* is the average of 40 female faces. In the matrix of faces, rows A–C show faces derived from three different faces; row D shows faces that are the means of Faces A–C. Over the matrix columns 1–5, faces deviate increasingly from the mean, by .04, in column 1, through .20, in column 5. For additional details, see the text.

of face space, which means that the geometric properties of the faces in D are equally well correlated with the geometric properties of each of the other faces, A, B, and C. Figure 2A shows the geometric arrangement of all 21 synthetic faces in a space of three orthogonal dimensions. To minimize potential influences from emotional expressions on the stimulus faces (Goren & Wilson, 2006; Isaacowitz, Wadlinger, Goren, & Wilson, 2006), our set of Wilson faces substituted constant generic shapes for features that would change shape or relative position as emotions were expressed (Ekman & Friesen, 1975).

Procedure. On each trial, a study set of three faces, the *study series*, was followed by a single *probe* face (**p**). The subjects judged whether **p** had been among the items in the study series. We use the term *target* to designate a **p** that had been in the study series and the term *lure* to designate a **p** that had not been in the study series. Correspondingly, we can designate any trial as either a target trial or a lure trial. Because the study series varied from trial to trial, the subjects were forced to base each *yes–no* recognition judgment on the items they had just seen.

Each study face was presented for 110 msec, with an interstimulus interval of 200 msec. The use of brief presentations was inspired by Wilson et al.'s (2002) use of this same duration in their studies of face discrimination and by the fact that fairly detailed processing of a face can be completed within the first 100 msec of viewing (Lehky, 2000).

A warning tone followed the study series. Then, after a 1,200-msec retention interval, a **p** face was presented for 110 msec. For each study list, **p** was chosen at random from the entire set of faces, with two constraints. First, on half of all trials, **p** was forced to replicate one of the items in the study set (on half of all the trials, **p** differed from all the study items). Second, when **p** matched one of the study items, it matched items in each serial position equally often. Distinctive tones following each response gave the subjects trialwise knowledge of results.

Although there were small differences in size from face to face, each was approximately 5.5º high × 3.8º wide. To eliminate the usefulness of Vernier-type cues, the vertical and horizontal position of each face was perturbed on each presentation by adding a pair of random displacements drawn from a uniform distribution with mean = 12.6 minarc and range = 2.1–23.1 minarc.

Sixty different stimulus series were used. A preliminary study with different subjects and many more stimulus series identified these 60 series as likely to span a wide range of recognition performance (Yotsumoto, Kahana, Wilson, & Sekuler, 2004). We reasoned that a wide range of performance would allow for the strongest test of NEMO. Of the 60 stimulus series, half consisted of target trials, in which *p* replicated one of three study items; the remaining lists consisted of lure trials, in which **p** replicated none of the study items.

The subjects participated in four 1-h sessions of 490 trials each. The first 10 trials of any session were treated as practice and were eliminated from data analysis; this left 32 replications for each stimulus series and subject. During testing, a subject sat with head supported by a chin-and-forehead rest, viewing the computer display binocularly from a distance of 114 cm. Trials were self-initiated.

Multidimensional scaling. To characterize the perceptual similarity structure of the synthetic faces (Lee, Byatt, & Rhodes, 2000), the subjects who would serve in the recognition memory experiment first took part in a study with nonmetric MDS. The data required for such scaling were generated using the method of triads (Ennis, Mullen, Frijters, & Tindall, 1989). On each trial, three faces were presented simultaneously, side by side, for 500 msec. Simultaneous presentation was used in order to minimize likely effects of

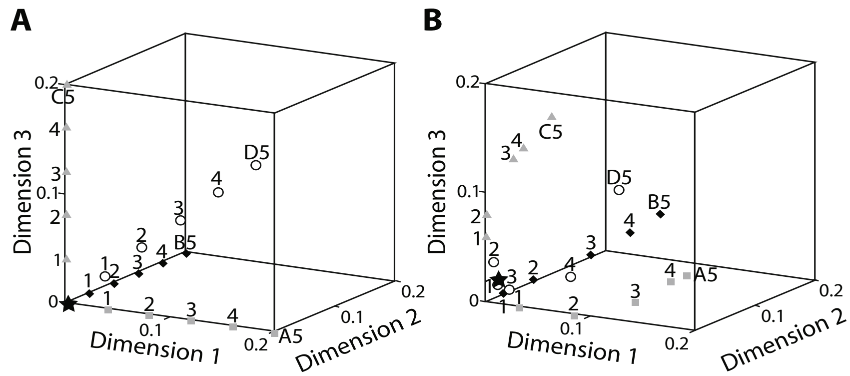

Figure 2. Representations of face stimuli in alternative 3-D spaces. In each panel, the star indicates the location of the mean face; squares, diamonds, circles, and triangles represent faces from Categories A, B, C, and D, respectively. The numbers 1–5 designate the faces' distance from the mean; the numbers correspond to the columns in Figure 1. (A) The 21 face stimuli arranged according to the Euclidean distances between the faces' physical descriptions—that is, the physical distance of each face from the mean. (B) Arrangement of the stimulus faces in a 3-D perceptual space, using the MDS solution to position each face (Experiment 2). The MDS space shown here has been Procrustes transformed to bring its dimensions into line with those of the space shown in panel A.

memory. From the set of three faces, the subjects chose the one face that seemed most different from the other two (Romney, Brewer, & Batchelder, 1993; Wexler & Romney, 1972). We did not specify the characteristic(s) on which similarity judgments should be based.

To minimize the possibility that Vernier-type cues might contribute to the dissimilarity judgments, each face's vertical position was randomly offset by a sample from a uniform distribution spanning ± 16.8 min. Each possible stimulus pair ($Face_1$, $Face_2$) was presented with every other possible stimulus—for example, $Face_3$. If stimulus $Face_3$ was selected as the stimulus most different from the others, then by default, the remaining stimuli, $Face_1$ and $Face_2$, were assumed to be the most similar to one another. A similarity matrix was constructed by counting the number of times that a stimulus pair (e.g., $Face_1$, $Face_2$) was designated as *similar* when placed in combination with various other stimuli (e.g., $Face_3 \ldots Face_{21}$).

To control the number of trials required for MDS, we used a balanced incomplete block design (Weller & Romney, 1988). For this design, we generated triads of faces (*blocks*), whose members were drawn from the complete set of 21 faces. This selection was constrained so that each of the 210 *pairs* of faces occurred in the context of 30 triads. This arrangement meant that the 30 trials whose triads included any particular pair of faces were likely to have different faces as their third member. The spatial displacements of the three faces were randomly determined anew for each trial.

Each subject participated in three 1-h sessions of 710 trials each. The first 10 trials of each session were treated as practice and were eliminated from our data analysis. The remaining 2,100 triadic comparisons per subject were converted into a dissimilarity matrix, which were processed by SPSS's ALSCAL and INDSCAL routines, using a Euclidean distance model.

Results

Recognition memory. Figure 3 shows the mean proportion correct for lure trials (arrow at right side of figure) and for target trials as a function of **p**'s serial position (three gray bars). An ANOVA showed that recognition on target trials across the three serial positions varied significantly [$F(2,14) = 24.43$, $p < .001$], and an a priori comparison revealed a significant effect of recency [$F(1,7) = 32.98$, $p < .01$]. For comparison, the stars in Figure 3 represent results with these same 60 stimulus series in another experiment with different subjects and many more stimulus series (Yotsumoto et al., 2004). The comparability of results from the two independent replications is gratifying. Note that performance for each serial position is significantly greater than chance; that is, each hit rate is significantly greater than the false alarm rate (paired two-tailed *t* tests: $p < .02$, .01, and .001 for Serial Positions 1, 2, and 3, respectively). So although performance with the first two serial positions is low, it is above chance.

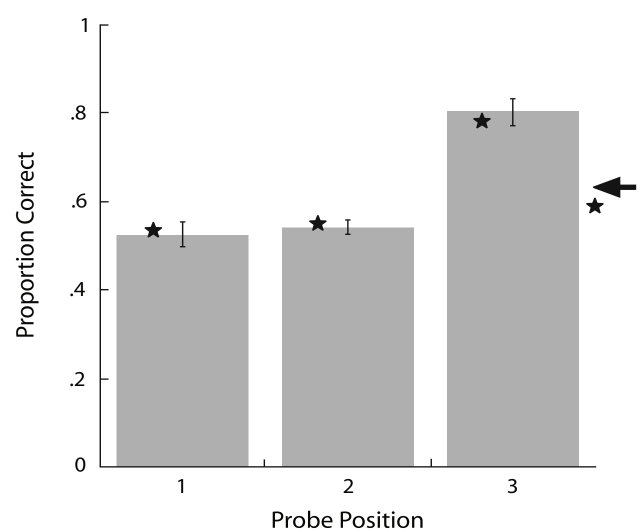

Figure 3. Proportions of correct recognition responses on target trials for probe matching various study items. Error bars represent ± 1 standard error of the mean, calculated using the correction suggested by Loftus and Masson (1994). The arrow to the right of the graph represents the mean proportion correct on lure trials. Each star shows the proportion correct generated by 29 different subjects who tested in the same conditions during an experiment described elsewhere (Yotsumoto, Kahana, Wilson, & Sekuler, 2004).

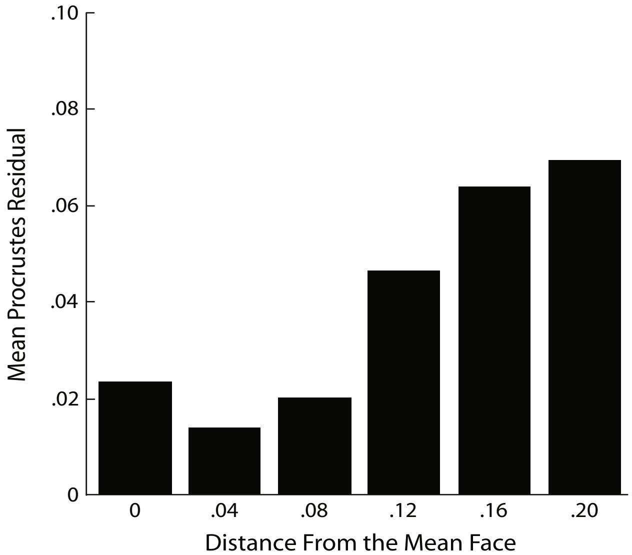

Figure 4. Mean distance between faces' representations in physical space and in perceptual (MDS) space, as a function of distance from the mean face.

Multidimensional scaling. MDS solutions were obtained for representations in one to six dimensions. Values of r^2, which represent the proportion of variance accounted for in the scaled data, increased as the number of dimensions varied from one to three but saturated thereafter. On the basis of these values and also on the dimensionality of the stimuli, subsequent modeling was based on the 3-D MDS solution. In that solution, Kruskal's stress measure was 0.26, and r^2 was .62. The 3-D solution generated by MDS is shown in Figure 2B. Each symbol represents one synthetic face, and the pairwise distances between symbols represent corresponding pairwise perceptual dissimilarities between the faces.

We used Procrustes analysis (Dryden & Mardia, 1998) to visualize relationships between this 3-D description of similarity space and the 3-D structure in which the faces were generated. The Procrustes analysis linearly transformed the matrix of values from the MDS solution to bring that matrix into best conformity with the matrix of pairwise distances in the faces' physical space. The outcome, shown in Figure 2B, was based on the Euclidean similarity transformations of translation, reflection, orthogonal rotation, and isotropic scaling of points in the MDS solution. If the perceptual representation of these synthetic faces were identical to their physical representation, the post-Procrustes MDS solution would be perfectly congruent with the representation in Figure 2A, in which Faces A, B, and C are orthogonal to each other and exemplars of Face D lie on the diagonal.

Clearly, although the transformed MDS solution does resemble the arrangement of the face stimuli themselves, residual differences remain between the perceptual space, as represented by MDS, and the physical space. After the Procrustes transformation, the sum of squared residual discrepancies between the physical representations and the transformed MDS representations was 0.26. To provide an intuition about the magnitude of this value, we used Monte Carlo methods to put this sum of squares into the same units as those used for the Euclidean physical space (Figure 2A). A series of Procrustes analyses were done on matrices in which the faces' physical coordinates were randomly perturbed to varying, known degrees, by the addition of independent, zero-mean, Gaussian random deviates to each of the three coordinates for each face. This operation was carried out 1,000 times for Gaussian distributions with different standard deviations. From the mean residual sums of squares associated with each value of standard deviation, we identified the random perturbation of face coordinates that produced the same residual sum of squares as had been obtained with the Procrustes transformation of the MDS solution. The standard deviation of the residual difference between the MDS solution and the original physical coordinates was equivalent to 6%–7%, which corresponds to ~1.5 × the separation of neighboring faces within any single category of faces, A . . . D.

To examine the residuals on a finer scale, the mean residuals between the MDS solution and the faces' physical coordinates were calculated and then sorted into bins according to the distance between faces in a study series and the mean of the 21 faces. These values are plotted in Figure 4. Note that the magnitude of the residuals grew with increasing distance from the mean face, confirming that perceptual and physical representations of faces were most discrepant for the more extreme faces in our set.

As a further comparison between the MDS and the physical representations of our 21 faces, we computed the vector angles between perceptual exemplars of A, B, and C. These vector angles are shown in Table 1. Faces just 4% away from the mean face were excluded from these calculations because, in MDS space, those faces clustered tightly around the mean face, which made angle measurements for those faces meaningless. The mean angles based on the 8%, 12%, and 16% data were 89º, 70º, and 93º, suggesting that the perceptual similarity space preserved much, but not all, of the orthogonality that had been built into the faces' original 3-D space. However, all of the angle estimates dropped when the .20 faces were included in the calculations, which confirms the demonstration in Figure 4 that these extreme faces show the largest perceptual deviations from the geometry of Figure 2A's space.

Model

We applied NEMO to the recognition memory data. As was mentioned before, NEMO departs from the classic summed-similarity models of item recognition (e.g.,

Table 1
Angles Between Perceptual Representations of Face Categories A–C

Face Distance	AB Angle	AC Angle	BC Angle
8%	125.7	65.3	116.1
12%	83.2	74.2	80.7
16%	57.4	71.5	83.4
20%	46.5	54.9	63.3
Mean for 8%–20%	78.2	66.5	85.9
Mean for 8%–16%	88.8	70.3	93.4

Table 2
Best-Fitting Parameter Values for NEMO's Fit to the Data

Parameter	Meaning	Physical	MDS
σ_1	Dimension_1 noise	0.033	0.032
σ_2	Dimension_2 noise	0.072	0.046
σ_3	Dimension_3 noise	0.045	0.046
α_1	Forgetting of 1st item	0.470	0.570
α_2	Forgetting of 2nd item	0.400	0.450
β	Interitem similarity	−0.530	−0.340
τ	Tuning function steepness	9.200	11.140
RMSD		0.123	0.101

McKinley & Nosofsky, 1996; Nosofsky, 1986) by allowing recognition judgments to be determined not only by the similarity between the probe, **p**, on one hand, and each study stimulus, on the other, but also by similarities among study items themselves. Given a series of L study items, $s_1 \ldots s_L$, and a probe, **p**, NEMO responds *yes* if

$$\underbrace{\sum_{i=1}^{L} \alpha\eta\left(\mathbf{p}, s_i + \varepsilon_i\right)}_{\text{Summed Probe Item Similarity}} + \beta \underbrace{\frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} \eta\left(s_i + \varepsilon_i, s_j + \varepsilon_j\right)}_{\text{Mean Within-List Homogeneity}} > C_L, \qquad (1)$$

where $\eta(\mathbf{p}, s_i)$ is the perceptual similarity between **p** and the ith study item (see Equation 2, below), $\boldsymbol{\varepsilon}$ is a vector representing the noise associated with each stimulus dimension, α_i is the weight given the ith study item, and C_L represents an optimal criterion for a series of L study items. To allow for the possibility that subjects' decision rule might incorporate within-list homogeneity, NEMO adds together (1) summed similarity and (2) within-list homogeneity, weighting the latter by a parameter β. If $\beta = 0$, the model reduces to a standard summed-similarity model (Nosofsky, 1986) with noisy item representations (Ennis, 1988) and a deterministic decision rule. If $\beta < 0$, when $s_1 \ldots s_L$ are similar to one another, a given lure becomes less tempting; that is, it attracts fewer *yes* responses. The opposite effect would accompany $\beta > 0$, that is, study items similar to one another would attract more *yes* responses.

In NEMO, similarity, $\eta(s_i, s_j)$, between item representations, s_i and s_j, is given by

$$\eta\left(s_i, s_j\right) = ae^{-\tau d\left(s_i, s_j\right)^c}, \qquad (2)$$

where d is the weighted distance between the two stimulus vectors and τ, c, and a jointly determine the form of the generalization gradient.

With similarity defined in either physical or perceptual (MDS) space, we ran parallel simulations of NEMO, with one set of simulations using each definition of similarity. Among parameters in Equation 2, we fixed $c = 1$, implementing a simple exponential generalization function, as suggested by previous empirical results (Kahana & Sekuler, 2002). To reduce the number of NEMO's free parameters further, we used an independent, empirical estimate of τ and a. We estimated similarity in physical space using data from a large-scale preliminary experiment (Yotsumoto et al., 2004) that generated an empirical approximation of the similarity tuning function in physical space.[1] The similarity tuning function's exponent and y-intercept, 9.20 and 0.84, were used for τ and a, respectively, in one set of model simulations. Then, reusing results from the preliminary experiment, we transformed interface distances according to values from the MDS analysis and fit a second exponential similarity function. This generated a similarity tuning function in perceptual space. The function's exponent and y-intercept, 11.14 and 0.91, were used as τ and a, respectively, in a second set of model simulations. Finally, we fixed one other parameter, setting NEMO's criterion to 0.5, which is the empirical value found in previous model fits (Kahana & Sekuler, 2002).

Simulations and application of model. We fit NEMO to the value of p(yes) obtained for each of the 60 different stimulus lists in Experiment 1. A genetic algorithm (Mitchell, 1996) found NEMO's best-fitting parameter set by minimizing the root-mean squared difference (RMSD) between observed and predicted recognition scores. The genetic algorithm allowed a population of 1,000 random parameter sets to evolve for 20 generations. At the end of every generation, each of the 500 least-fit parameter sets was replaced with a new parameter set, which randomly drew each of its parameter values from one of the nonreplaced 500 best-fitting parameter sets. Then the nonreplaced 500 best-fitting parameter sets were mutated by a single Gaussian parameter change with a standard deviation of 30% of a parameter's range. Finally, to produce an estimate of RMSD, each parameter set ran for 1,000 simulated trials for each stimulus list.

Results of model simulations. We fit the subjects' average performance twice, expressing NEMO's interface similarity values, τ, a, and $d(s_i, s_j)$, either as *physical* distances between faces or as values from the MDS descriptions of subjects' *perceptual* space. Table 2 gives the best-fitting model parameters for NEMO derived from the genetic algorithm. The "Physical" column shows the best parameters using stimulus distances in physical space, and the "MDS" column shows the best parameters using stimulus distances in MDS perceptual space. The first three parameters, σ_1, σ_2, and σ_3, are the variances of noise distributions—one for each dimension of the 3-D, perceptual space. σ_1, σ_2, and σ_3 correspond to Dimension 1, 2, and 3 in Figure 2. The next two parameters, α_1 and α_2, represent forgetting for the first and second items in a study series, respectively. (For the last item in a study series, α_3 was set to one.)

As was explained earlier, β represents the contribution of within-list homogeneity. Its negative sign for both simulations indicates that when study items were similar to one another, NEMO became more conservative, decreasing any tendency to treat a lure as a target. Note, finally, that the RMSD associated with the perception-based fit (0.101) is smaller than the RMSD associated with the fit based on the faces' physical representation (0.123).

Each of the best-fitting parameter sets was used to generate NEMO's predictions for each of the 60 lists. Because NEMO has three nondeterministic noise parameters, running the model with any single trio of noise samples

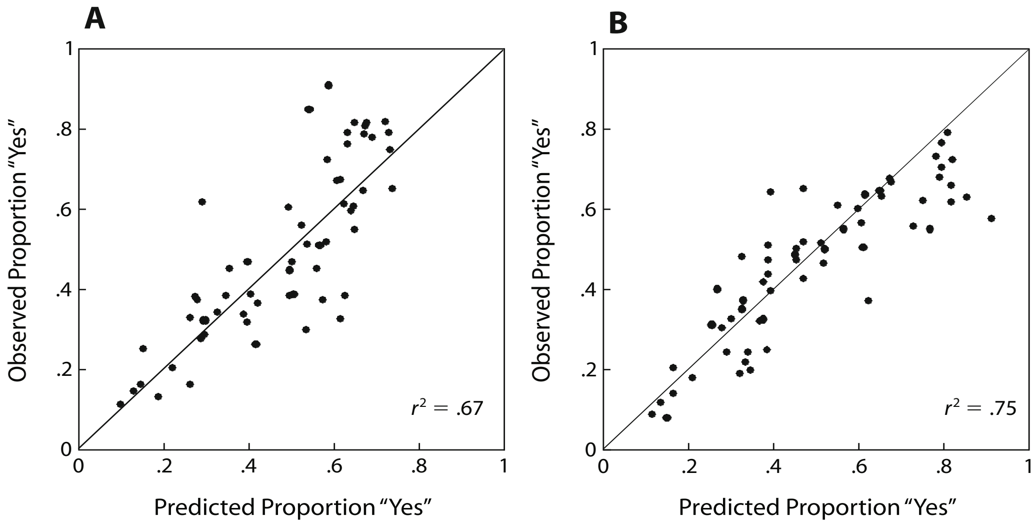

Figure 5. Proportions of *yes* responses plotted against predictions from NEMO. (A) Interstimulus distances used in NEMO were taken from physical descriptions of the stimuli. (B) Interstimulus distances used in NEMO were taken from the multidimensional scaling solution.

$\sigma_1 \ldots \sigma_3$ could not produce a singular, exact set of predictions. Therefore, NEMO was used to simulate 1,000 trials for each of the 60 study–**p** lists, with new, independent random noise samples drawn for each trial. From these 1,000 trials, we obtained the predicted proportion of *yes* responses for each list. Figure 5 shows, for each list, the relationship between the predicted and the observed proportions of *yes* responses. The predictions made with similarity defined by the physical representations of face stimuli are shown in Figure 5A; predictions based on the MDS solution are shown in Figure 5B. NEMO produced a better account of the data when it incorporated perceptual similarity among faces. Physical and perceptual similarities produced $r^2 = .68$ and .76, respectively. To gauge the reliability of this difference in r^2, we repeated the simulation an additional 1,000 times for each list, again drawing new independent noise samples for each trial and calculating r^2 anew after every new 1,000 trials. For simulations based on physical similarity between faces, the mean $r^2 = .6838$ ($SD = .0109$); for simulations based on MDS-defined similarity, the mean $r^2 = .7573$ ($SD = .0090$). The difference between these mean values of r^2 is highly significant ($z = 3.43$, $p < .001$).

The reliability of the MDS solutions. To assess the consistency of the subjects' triadic judgments, we computed two different mean MDS solutions. One solution was based on the subjects' dissimilarity judgments on all the odd-numbered trials (i.e., first, third, fifth, etc.); the second solution was based on judgments from all the even-numbered trials (i.e., second, fourth, sixth, etc.).[2] For each 3-D solution, the Euclidean distances between all the face pairs were taken, and the correlation was calculated between pairwise distances from odd trials and pairwise distances from even trials. The results are shown as a scatterplot in Figure 6. Despite the 50% reduction in the number of judgments on which each MDS solution was based, the pair of MDS solutions produced by this process had a relatively strong correlation, with $r^2 = .79$, which we take as confirmation that the triadic comparisons produced reliable measures of similarity.

As was noted earlier, NEMO's predictions were more accurate when psychophysical (MDS) rather than purely physical similarities were taken account of, but those predictions had a number of clear outliers. These were stimulus series on which the model failed badly—that is, deviated by .20 or more from the predicted p(yes). To identify the origin of these failures, we examined the makeup of these series. Of the five outliers, three contained two or

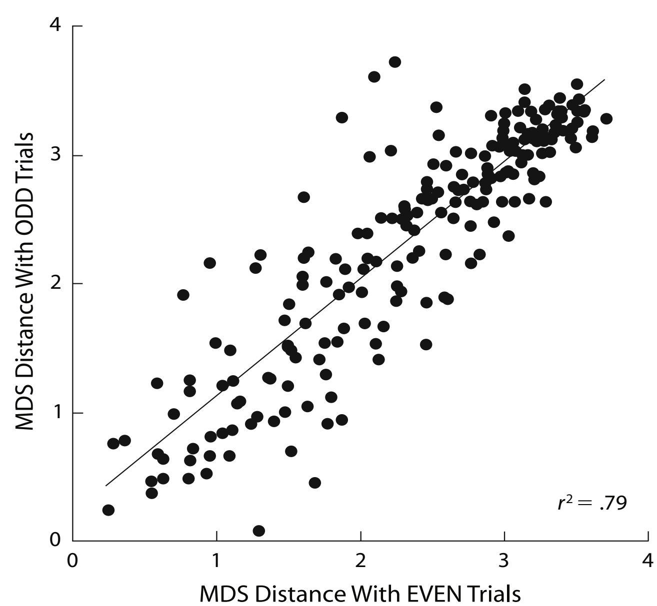

Figure 6. Distance between items in the multidimensional scaling (MDS) solution calculated using even trials and then calculated using odd trials.

more faces that deviated from the mean face by the largest value possible—namely, .20. To this outcome, Monte Carlo simulation assigned a probability $.02 < p < .01$. So, faces with the greatest deformation, relative to the mean face, produced the largest errors in NEMO's predictions. As is shown in Figure 4 and in Table 1, these extreme faces deviated appreciably from the orthogonal physical representations of the faces. These deviations might have arisen from perceptual transformations associated with the faces' "strangeness," which effectively introduced an additional perceptual dimension limited to the extreme faces. Because this added dimension was associated with only a subset of the faces, it was not represented fully in the MDS solution.

Discussion

NEMO's account of the recognition results gives further support to the idea that recognition decisions depend upon both summed similarity and within-list homogeneity. Discrepancies between the faces' representation in physical space (Figure 2A) and their perceptual representation (Figure 2B) advantaged model fits that were based on perceptual, rather than physical, similarities between faces. The perceptual representation was based on MDS, which requires as input a matrix of similarity or dissimilarity judgments. Researchers have taken various approaches to generate such matrices. For example, the input matrix has been generated by asking subjects to rate the distinctiveness of items presented one at a time (Lee et al., 2000; Valentine, 1991; Valentine & Bruce, 1986; Valentine & Endo, 1992) or to rate numerically the similarity of items presented in pairs (Johnstone & Williams, 1997; Nosofsky, 1991; Peters et al., 2003). We took a different approach, using triadic comparisons to produce the input matrix for MDS. We chose this method, in part, for its efficiency in generating many different comparisons per pair of faces and, in part, because the task resembled the recognition memory task we used. For one thing, by encouraging subjects to distinguish among members of the briefly presented triad, the task was likely to engage the rapid and sometime subtle distinctions required in the recognition judgments. At the same time, variation in the triad's constituents from one presentation to another mimicked the trialwise variation among study series.

Torgerson (1958) described other variants of the method of triadic comparisons in which subjects had to make multiple, explicit pairwise judgments per trial. Note that the single explicit judgment required on each trial in our application implies that subjects have made one or more pairwise comparisons, although such comparisons are not made explicit. Letting the stimuli in the triad be i, j, and k, our subjects' identification of one item as most dissimilar could reflect evaluations of inequalities among $|i - j|$, $|i - k|$, and $|j - k|$. Of course, when subjects are not forced to make such evaluations explicit, one cannot rule out the possibility that, particularly with time pressures, subjects might sometimes make fewer than all pairwise comparisons.

EXPERIMENT 2

The simulations applied to the data from Experiment 1 revealed that visual memory performance could be well predicted by a model that takes account of both summed similarity and within-list homogeneity. Because the within-list homogeneity term is a novel addition to the summed-similarity framework, we sought an additional, direct, model-free demonstration that within-list homogeneity was actually important in face recognition memory. Therefore, we designed stimulus series in which variation in both summed similarity and within-list homogeneity were controlled. We expected that the responses produced by various combinations of the two factors would directly demonstrate the contribution of each factor, without the mediation of a computational model.

Method

Apparatus, Stimuli, and Procedure. The apparatus and stimuli were the same as those in Experiment 1, except that multiple subjects were tested simultaneously, using computers in a classroom cluster. Although the subjects did not use chinrests, they were encouraged to maintain a constant viewing distance of approximately 57 cm from their computers. The procedure was the same as that in Experiment 1, except that the three study faces for each trial were forced to come from three different categories of faces, A . . . D (as shown in Figure 1). Study lists were first generated randomly. Because of the upper limit on possible pairwise, physical differences between the faces that we used, differences among randomly selected faces tended to be small, which produced skewed distributions of summed similarity and within-list homogeneity. Moreover, the randomly generated lists produced nonzero covariance between summed similarity and within-list homogeneity. To test wider ranges of both types of similarity independently, summed similarity and within-list homogeneity had to be distributed uniformly. To generate series meeting this criterion, the distributions of two kinds of similarity were calculated after random generation of a set of series, and existing stimulus series were replaced by newly generated ones until we had a set of study series that satisfied the distribution requirement.

Subjects. Twenty-nine Brandeis undergraduates participated as part of a course requirement; during the session, each subject took part in 436 trials. All the subjects were naive as to the experimental purpose, and none had taken part in our other experiments.

Results and Discussion

For each series on which p was a lure, we used the faces' physical coordinates to calculate summed similarity between p and all the study items and the within-list homogeneity. For this purpose, similarity was assumed to be monotonic with Euclidean distance in the physical space represented in Figure 2A. We sorted the trials into the cells of a 3×3 matrix, whose rows represented three levels of summed similarity and whose columns represented three levels of within-list homogeneity. At the end of the sorting, each cell of the 3×3 matrix contained 24 trials per subject. The proportion of *yes* responses was calculated separately for each of the nine cells in the matrix; these values are plotted in Figure 7. The parameter of the family of curves is within-list homogeneity. The proportion of *yes* responses increased with summed similarity but decreased with growth in within-list homogeneity. A repeated measures ANOVA showed that both these effects were statistically significant [$F(2,56) = 85.6$, $p < .01$, and $F(2,56) = 7.08$, $p < .05$, for summed similarity and within-list homogeneity, respectively]. The interaction was not significant [$F(4,112) = 1.16$, $p = .33$].

Note that the directions of the two effects observed here reproduced the corresponding effects seen in simulations with NEMO for Experiment 1. Of particular interest was the direct confirmation that within-list homogeneity and summed similarity operate in opposed directions to influence recognition judgments, just as NEMO demonstrated they did.

GENERAL DISCUSSION

Physical Coordinates Versus MDS Solutions

NEMO's account of visual recognition memory performance for face stimuli was improved when the model incorporated faces' perceptual similarity, rather than purely physical coordinates. It is important to note that with either perceptual or physical representations of similarity, NEMO had the same number of free parameters; therefore, this difference did not result from a difference in model complexity.

We should note that not every related study has demonstrated an advantage from describing stimuli in perceptual (MDS) terms, rather than physical ones. Peters et al. (2003) found that the performance of categorization models was either unchanged or even slightly diminished when face stimuli were described using MDS, rather than a native physical metric. In their study, subjects were trained to categorize stimuli including schematic Brunswik–Reiter faces (1937) and slightly more elaborate cartoon faces. These faces were defined in 4-D physical spaces, which the subjects learned to bifurcate using a simple, linear separable criterion. After learning the category membership of various exemplars, the subjects' categorization was tested with mixtures of previously seen faces and new ones. Peters et al.'s modeling results favored the proposition that the subjects stored a sparse, abstracted representation of category properties, rather than the characteristics of individual exemplars. This demonstration of longer term learning of stable category properties is quite different from the case in our experiments on episodic recognition, where the subjects seem to have stored individual exemplars, at least for the brief duration of a single trial. In categorization tasks such as Peters et al.'s, subjects can learn abstract rules over the course of their experience with many exemplars. However, such a strategy would not work in episodic recognition tasks in which targets and lures are drawn randomly from a common stimulus space and in which no simple rule could work in the face of trial-to-trial variation in the study and test items. Of course, if some simplifying consistent bias were introduced into a recognition experiment so that lures were always drawn from one category (say, Wilson faces from Class A) and targets were always drawn from a different category (say, Wilson faces from Class B), subjects undoubtedly would eventually learn the rule and be able to ignore the study items. Clearly, though, such an experiment would no longer be an experiment on episodic recognition.

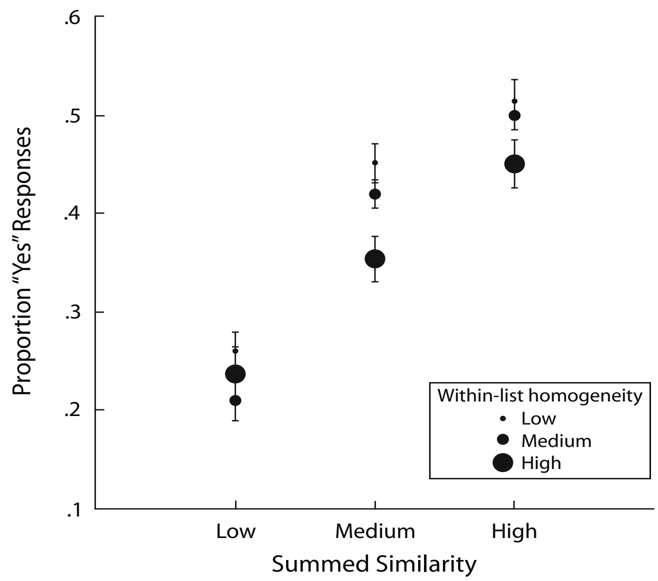

Figure 7. The proportion of *yes* responses on lure trials as a function of summed similarity. These false alarm rates are plotted separately for three different levels of within-list homogeneity. The diameter of each filled circle signifies the magnitude of within-list homogeneity. Error bars represent ±1 standard error of the mean, corrected for within-subjects variability (Loftus & Masson, 1994).

Differences From Compound Gratings

One of our purposes was to investigate short-term visual memory with higher dimensional stimuli, particularly by applying NEMO to the recognition of synthetic faces. The best-fitting parameters obtained here preserved the general characteristics observed in Kahana and Sekuler's (2002) experiments with memory for compound gratings. For example, in both experiments, recency effects were captured by values of α, and the empirical effects of within-list homogeneity were reflected in significantly negative values of β. Moreover, despite the fact that τ values were derived in different ways for recognition of gratings and of faces, the resulting τ values were close between the two experiments (8.8 and 10.7 with compound gratings, 9.20 and 11.14 with synthetic faces).[3] It seems, then, that similar similarity-distance functions operate for both low-dimensional (gratings) and high-dimensional (synthetic faces) stimuli.

However, memory for synthetic faces may differ in an important way from the memory for compound gratings. Even though we found a recency effect with synthetic faces, as well as with gratings, attributes of the recency effect observed in this study differed from those found with gratings. With gratings, the recency effect persisted across the entire list, with performance increasing systematically from the least to the most recently presented item (Kahana & Sekuler, 2002). Here, though, the serial positions preceding the last one produced essentially equivalent performance, although, as was mentioned earlier, performance at all serial positions was better than the chance level defined by the false alarm rate. This suggests that higher dimensional stimuli, instead of promoting a gradual forgetting of previously seen items, allow the last item seen to diminish the memory for all previously seen items equally. This resembles a result reported previously (Phillips, 1974, 1983; Phillips & Christie, 1977).

Because procedures differed between the studies with grating and face stimuli, we must be cautious in attributing various discrepancies in the results to differences in stimulus dimensionality alone. But we do believe that parallel studies of recognition memory for gratings, faces, and other high-dimensional stimuli, as well as studies of comparable stimuli from other sensory modalities (Visscher, Kaplan, Kahana, & Sekuler, 2006), will ultimately help us understand how the number and structure of perceptual dimensions that make up stimuli contribute to human short-term recognition memory.

We should note at least one possible application of the present results, to the validity of police lineups. When a witness views a simultaneous lineup comprising several individuals, the homogeneity of those individuals is an analogue to the within-list homogeneity whose potency was demonstrated in Experiments 1 and 2. Extrapolating from our results, one would expect that within-list homogeneity in a lineup would strongly affect a potential witness' recognition response. Recently, law enforcement officials in the United States have been encouraged to substitute sequential lineups for the traditional simultaneous procedure (Turtle, Lindsay, & Wells, 2003). Advocates of this procedure have claimed that sequential lineups somehow enhance discriminability and, thereby, promote accuracy. In sequential lineups, witnesses view one lineup member at a time and decide whether or not that person is the perpetrator, prior to viewing the next lineup member. Here, witnesses make a *yes–no* judgment after viewing each single person/face, one at a time. This is meant to minimize false alarms, by discouraging relative judgments that can contaminate simultaneous lineups (Lindsay et al., 1991; Steblay, Dysart, Fulero, & Lindsay, 2001). However, Gronlund (2004) has shown that the beneficial effect of sequential lineups arises not from heightened discriminability, but from a change in criterion—with a sequential lineup promoting a more conservative criterion. Whatever their benefit, though, sequential lineups would be immune to influences from within-list homogeneity only if there were zero carryover of memory from one face/lineup member to another, an assumption that begs to be evaluated.

Finally, in all the experiments we have reported here, possible contamination of face recognition by emotional cues was intentionally avoided. In fact, the set of Wilson faces that we used substituted constant generic shapes for features that would change shape as emotions were expressed. In addition to equation on basic dimensions such as contrast and mean luminance, the consistent neutral expression of our faces ruled out emotion as an aid to recognition memory. Given emotional expression's known role in recognition (e.g., Gallegos & Tranel, 2005; Johansson, Mecklinger, & Treese, 2004; Kaufmann & Schweinberger, 2004), it is worth noting that systematic variation in the position and/or orientation of eyes, mouth, and brows (Ekman & Friesen, 1975) can generate Wilson faces that express distinct, easily identified emotions, and in varying degree (e.g., Isaacowitz et al., 2006). We plan to evaluate how the presence of emotion signals of varying, calibrated strength might combine with other facial information to influence short-term face recognition.

AUTHOR NOTE

The authors acknowledge support from AFOSR F49620-03-1-0376 and National Institutes of Health Grants MH55687 and EY002158. Y.Y. is now at the Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown. Correspondence concerning this article may be addressed to R. Sekuler, Volen Center for Complex Systems, Brandeis University, Waltham, MA 02454 (e-mail: sekuler@brandeis.edu).

REFERENCES

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, **5**, 204-210.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 251-301). Mahwah, NJ: Erlbaum.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 443-446.

Brunswik, E., & Reiter, L. (1937). Eindruckscharaktere schematisierter Gesichter. *Zeitschrift für Psychologie*, **142**, 67-134.

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.

Diamantaras, K. I., & Kung, S. Y. (1996). *Principal component neural networks*. New York: Wiley.

Dryden, I. L., & Mardia, V. (1998). *Statistical shape analysis*. New York: Wiley.

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, **41**, 713-720.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice-Hall.

Ennis, D. M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, **117**, 408-411.

Ennis, D. M., Mullen, K., Frijters, J. E. R., & Tindall, J. (1989). Decision conflicts: Within-trial resampling in Richardson's method of triads. *British Journal of Mathematical & Statistical Psychology*, **42**, 265-269.

Gallegos, D. R., & Tranel, D. (2005). Positive facial affect facilitates identification of famous faces. *Brain & Language*, **93**, 338-348.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, **3**, 191-197.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, **402**, 176-178.

Goldstein, A. G., & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics*, **9**, 237-241.

Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, **46**, 1253-1262.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, **7**, 555-562.

Gronlund, S. D. (2004). Sequential lineups: Shift in criterion or decision strategy? *Journal of Applied Psychology*, **89**, 362-368.

Hole, G. J. (1996). Decay and interference effects in visuospatial short-term memory. *Perception*, **25**, 53-64.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, **33**, 36-67.

Hwang, G., Jacobs, J., Geller, A., Danker, J., Sekuler, R., & Kahana, M. J. (2005). EEG correlates of subvocal rehearsal in working memory. *Behavioral & Brain Function*, **1**, 20.

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

Johansson, M., Mecklinger, A., & Treese, A.-C. (2004). Recognition memory for emotional and neutral faces: An event-related potential study. *Journal of Cognitive Neuroscience*, **16**, 1840-1853.

Johnstone, A. B., & Williams, C. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, **4**, 59-67.

Joseph, J. E., & Gathers, A. D. (2002). Natural and manufactured objects activate the fusiform face area. *NeuroReport*, **13**, 935-938.

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

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

Kaufmann, J. M., & Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception*, **33**, 399-408.

Klein, I., Paradis, A. L., Poline, J. B., Kosslyn, S. M., & Le Bihan, D. (2000). Transient activity in the human calcarine cortex during visual-mental imagery: An event-related fMRI study. *Journal of Cognitive Neuroscience*, **12**(Suppl. 2), 15-23.

Kosslyn, S., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, **378**, 496-498.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space hypothesis. *Psychological Science*, **11**, 379-385.

Lehky, S. R. (2000). Fine discrimination of faces can be performed rapidly. *Journal of Cognitive Neuroscience*, **12**, 848-855.

Lindsay, R. C., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, **76**, 796-802.

Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research*, **45**, 2287-2297.

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

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, **62**, 81-92.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 294-317.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, **39**, 3824-3833.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-394). Hillsdale, NJ: Erlbaum.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). Designing a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, **2**, 187-199.

Peters, R. J., Gabbiani, F., & Koch, C. (2003). Human visual object categorization can be described by models with low memory capacity. *Vision Research*, **43**, 2265-2280.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, **16**, 283-290.

Phillips, W. A. (1983). Short-term visual memory. *Philosophical Transactions of the Royal Society of London: Series B*, **302**, 295-309.

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, **29**, 117-133.

Principle, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and adaptive systems*. New York: Wiley.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London: Series B*, **271**, S448-S450.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, **4**, 28-34.

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, **32**, 285-293.

Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, **14**, 187-198.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law & Human Behavior*, **25**, 459-473.

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

Turtle, J., Lindsay, R. C., & Wells, G. L. (2003). Best practice recommendations for eyewitness evidence procedures: New ideas for the oldest way to solve a case. *Canadian Journal of Police & Security Services*, **1**, 5-18.

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). The effects of distinctiveness in recognising and classifying faces. *Perception*, **15**, 525-535.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology*, **44A**, 671-703.

Visscher, K., Kaplan, E., Kahana, M. J., & Sekuler, R. (2006). Auditory short-term memory behaves like visual short-term memory. *Public Library of Science Biology*, **5**, e56.

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

Wexler, K. N., & Romney, A. K. (1972). Individual variations in cognitive structures. In A. K. Romney, R. M. Shepard, & M. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (2nd ed., pp. 73-92). New York: Seminar Press.

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., Wilson, H. R., & Sekuler, R. (2004). *Preliminary studies of recognition memory for synthetic faces* (Tech. Rep. 2004-3). Waltham, MA: Brandeis University, Visual Cognition Laboratory, Volen Center for Complex Systems.

Zhou, F., Kahana, M. J., & Sekuler, R. (2004). Short-term episodic memory for visual textures: A roving probe gathers some memory. *Psychological Science*, **153**, 112-118.

NOTES

1. The similarity tuning function was based on errors in *same–different* judgments with just a single study item, which minimized memory load, and with timing identical to that used here.

2. The balanced incomplete block design forced us to take this indirect approach. Differences in the makeup of triads on successive trials prevented us from comparing judgments themselves. As a result, our comparisons had to be mediated via MDS.

3. We ran simulations with various τ values and found that differences of this size did not appreciably alter the model's resulting RMSD. So differences in the similarity-distance functions were not critical to the success of model fits.

(*Continued on next page*)

APPENDIX

Wilson et al. (2002) introduced a method for generating synthetic faces that are well suited for model-driven research on various topics, including visual memory. In their scheme, individual synthetic faces are derived from grayscale face photographs by digitizing 37 key points: 14 points defining head shape, 9 points for the hairline, 4 points for eye locations, 4 points for nose length and width, 5 points defining the mouth and lips, and 1 point for brow height. Synthetic faces are then reconstructed from these 37 measurements and bandpass filtered with a 2.0-octave-wide difference of Gaussian filter with a peak frequency of 10.0 cycles per face width (Wilson et al., 2002). Several studies have shown that such filtering preserves frequencies needed for face recognition (Gold et al., 1999; Näsänen, 1999).

By design, synthetic faces eliminate textures, such as skin, hair, wrinkles, and so forth, and focus instead on the geometric characteristics of faces. However, this raises the important question of whether synthetic faces are sufficiently accurate representations of their original faces to be useful in psychophysical experimentation. This question has been answered by requiring observers to identify the grayscale photograph from which a synthetic face was derived in a four-alternative forced choice experiment. The mean across 5 observers was 97.4% correct in matching between front view synthetic faces and photographs, and even for matching between 20 side view photographs and front view synthetic faces (or vice versa), performance averaged 90.7% correct (Wilson et al., 2002). Because chance performance is 25% in these experiments, these data clearly demonstrate that synthetic faces capture salient aspects of individual face geometry. Furthermore, the database captures known face gender differences: Synthetic female faces have significantly smaller heads, rounder chins, thicker lips, and higher eyebrows than do males. Finally, fMRI signals from the fusiform face area show that synthetic faces produce BOLD activation that is nearly as large as the original grayscale faces from which they are derived (Loffler, Yourganov, et al., 2005). Although the Wilson faces are relatively simple geometrically, they still convey sufficient information to characterize individual faces. This essential individuality, which is important in episodic memory, is absent from some commonly used face stimuli, such as Brunswik faces (Brunswik & Reiter, 1937; Peters et al., 2003; Sigala et al., 2002).

(Manuscript received December 17, 2005;
revision accepted for publication July 12, 2006.)