Cognitive Science: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653634

Modeling Recognition Memory Using the Similarity Structure of Natural Input

Joyca P. W. Lacroix [a]; Jaap M. J. Murre [b]; Eric O. Postma [a]; H. Jaap van den Herik [a]
[a] IKAT/Department of Computer Science, Universiteit Maastricht, The Netherlands. [b] IKAT/Department of Computer Science, Universiteit Maastricht, The Netherlands; Department of Psychology, University of Amsterdam, The Netherlands.

Online Publication Date: 01 January 2006

To cite this Article Lacroix, Joyca P. W., Murre, Jaap M. J., Postma, Eric O. and Herik, H. Jaap van den(2006)'Modeling Recognition Memory Using the Similarity Structure of Natural Input',Cognitive Science: A Multidisciplinary Journal,30:1,121 — 145
To link to this Article: DOI: 10.1207/s15516709cog0000_48
URL: http://dx.doi.org/10.1207/s15516709cog0000_48

Cognitive Science 30 (2006) 121–145

Modeling Recognition Memory Using the Similarity Structure of Natural Input

Joyca P. W. Lacroix[a], Jaap M. J. Murre[a, b], Eric O. Postma[a], H. Jaap van den Herik[a]

[a]*IKAT/Department of Computer Science, Universiteit Maastricht, The Netherlands*
[b]*Department of Psychology, University of Amsterdam, The Netherlands*

Received 3 August 2004; received in revised form 11 July 2005; accepted 5 August 2005

Abstract

The natural input memory (NIM) model is a new model for recognition memory that operates on natural visual input. A biologically informed perceptual preprocessing method takes local samples (eye fixations) from a natural image and translates these into a feature-vector representation. During recognition, the model compares incoming preprocessed natural input to stored representations. By complementing the recognition memory process with a perceptual front end, the NIM model is able to make predictions about memorability based directly on individual natural stimuli. We demonstrate that the NIM model is able to simulate experimentally obtained similarity ratings and recognition memory for individual stimuli (i.e., face images).

Keywords: Psychology; Memory; Perception; Representation; Artificial intelligence; Computer science; Computer simulation

1. Introduction

Many computational memory models represent an item as a vector of abstract features, that is, a point in a multidimensional space. The feature values are usually generated using a mathematically defined distribution. For instance, when Shiffrin and Steyvers (1997) proposed their Retrieving Effectively from Memory (REM) model, they used the geometric distribution to select the feature-vector values. Other models that followed this approach include the SAM model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), the TODAM model (Murdock, 1982), the CHARM model (Eich, 1982, 1985), the Matrix model (Pike, 1984), the MINERVA2 model (Hintzman, 1986), the model of differentiation (McClelland & Chappell, 1998), and the BCDMEM model (Dennis & Humphreys, 2001). Because these computational

Correspondence should be addressed to Joyca P. W. Lacroix, IKAT/Department of Computer Science, Universiteit Maastricht, P.O. Box 616, 6200 MD, Maastricht, The Netherlands. E-mail: j.lacroix@cs.unimaas.nl

models operate on artificial vector representations, they are unable to make predictions for natural stimuli in behavioral memory experiments.

The similarity structure of natural stimuli can be represented in any type of space that is based on the compactness hypothesis (Arkadev & Braverman, 1966). This hypothesis states that similar objects in the real world are close in their representations and that there is no ground for any generalization on representations that do not obey this demand. In accordance with this hypothesis, several researchers developed so-called similarity spaces, which represent similar stimuli in close proximity of one another. This was done for conceptually based stimuli, such as words (e.g., Landauer & Dumais, 1997; Steyvers, Shiffrin, & Nelson, 2004), as well as for perceptually based stimuli, such as objects and faces (e.g., Busey, 1998; Edelman, 1995, 1998; Nosofsky, 1986; Valentine, 1991). For the construction of perceptually based similarity spaces, an analysis of human similarity judgments is often used. This leads to a psychologically plausible representation space, in which the perceived similarity between two stimuli directly translates into the distance between their representations. However, a severe limitation is that representations are not derived directly from the perceptual features of the stimulus. Instead, the representations are established *a posteriori* in terms of the relative distances to the stimuli considered in the similarity analysis. In the following, we propose a recognition memory model that operates on natural images: the natural input memory (NIM) model. The NIM model is a recognition version of the generalized context model (GCM; e.g., Nosofsky, 1986, 1987) extended with a perceptual front end. When presented with a natural image, the NIM model employs a biologically informed perceptual preprocessing method that translates the image into a similarity-space representation. In our simulations, we use a set of digitized gray-scale images of male human faces without glasses (e.g., Busey & Tunnicliff, 1999).

This article is organized as follows. In the next section, we introduce and describe the NIM model. Subsequently, we validate the NIM model with individual natural stimuli in two tasks. First, the NIM model is tested on a similarity-rating task, to assess the psychological plausibility of the similarity space that the model builds from the natural input. This is followed by a validation of the NIM model on a face-recognition task. Finally, we discuss the main difference between the NIM model and existing models, that is, the perceptual front end, and describe possible extensions of the model.

2. The NIM model

The NIM model encompasses the following two stages:

1. A perceptual preprocessing stage that translates a natural image into feature vectors.
2. A memory stage composed of two processes:
   a. a storage process that stores feature vectors in a straightforward manner;
   b. a recognition process that compares feature vectors of the image to be recognized with previously stored feature vectors.

Fig. 1 presents a schematic diagram of the NIM model. The face image is an example of a natural image. The left and right side of the diagram correspond to the two stages of the NIM model: the perceptual preprocessing stage (left) and the memory stage (right). The preprocess-

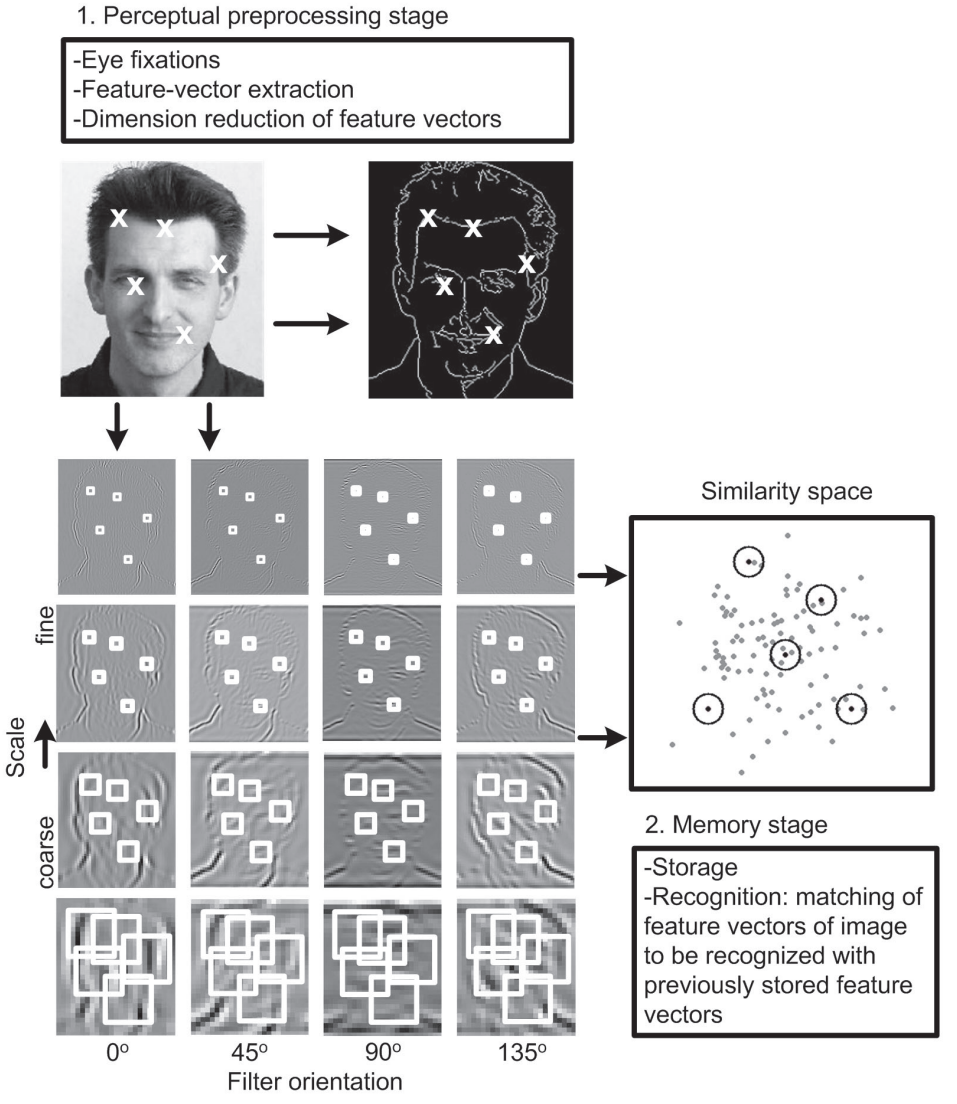

Fig. 1. The NIM model.

ing stage selects eye-fixation locations and extracts perceptual input (i.e., a feature vector) at each eye-fixation location. Then, it reduces the dimensionality of the feature vectors (not shown in the diagram), which leads to a low-dimensional feature-vector representation. The low-dimensional feature-vector representation forms the input of the memory stage (other than that there are no interactions between the preprocessing and memory stages). During storage, the memory stage stores the feature-vector representation. During recognition, the memory stage compares the feature-vector representation with previously stored representations. In the following section, the perceptual preprocessing stage and the memory stage will be explained in more detail.

2.1. The perceptual preprocessing stage

The perceptual preprocessing stage is inspired by biological knowledge about the processing of information in the human visual system (e.g., Hubel, 1988; Palmer, 1999) and by certain computational considerations (e.g., Bellman, 1961; Bishop, 1995; Edelman & Intrator, 1997). These are discussed in this section. In addition, we provide some relevant implementation details.

2.1.1. Biological inspiration and computational considerations

Early visual processing in the brain leads to the activation of millions of optic nerve cells (Palmer, 1999). The nerve-cell activations may be conceived as a high-dimensional vector. The high dimensionality enables the representation of a large amount of information without suffering from interference (Rao & Ballard, 1995), but it also hampers the memory performance, as the number of examples that are necessary for a reliable generalization performance grows exponentially with the number of dimensions. This phenomenon is known as the *curse of dimensionality* (Bellman, 1961; Bishop, 1995; Edelman & Intrator, 1997). In coping with this phenomenon, subsequent stages in the visual system are assumed to reduce the dimensionality of the high-dimensional input (e.g., Barlow, 1989; Hubel, 1988; Tenenbaum, Silva, & Langford, 2000). This assumption is supported by findings of Edelman and Intrator, who showed that the human visual system can be well described as extracting the low-dimensional structure of high-dimensional visual input.

In the NIM model, dimension reduction of high-dimensional natural input is achieved by a biologically informed feature-vector extraction method (Freeman & Adelson, 1991). The feature-vector extraction method employed by the NIM model operates directly on a high-dimensional digitized natural image of a face. The image has a high dimensionality because it is treated as a vector, the elements of which are the constituent pixel values. Motivated by eye fixations in human vision, the feature-vector extraction method takes local samples from selected locations in the image. To emphasize the parallel with human vision, we refer to these samples as *fixations.* Human fixations tend to cluster at or near contours (e.g., Norman, Phillips, & Ross, 2001; Yarbus, 1967). The preference for contours can be explained in terms of the principle of maximizing information (e.g., Petrov & Zhaoping, 2003; Wainwright, 1999). Natural images are characterized by a high degree of redundancy, mainly caused by intensity correlations among adjacent pixels. At contours, intensity correlations between adjacent pixels are low, which makes contours highly informative. As in the human eye, the feature-vector extraction method employed by the NIM model places fixations along the contours of the image.

For each fixation, features (i.e., a feature vector) are extracted from the image area centered at the fixation location. Feature-vector extraction in the NIM model mimics the visual processing in area V1. The responses of neurons in V1 are modeled by a multiscale wavelet decomposition (described in detail in the next section). Multiscale wavelet decomposition models are biologically plausible (e.g., Lee, 1998; Palmeri & Gauthier, 2004). Moreover, several studies showed that distances between representations in the similarity space that results from preprocessing input with the biologically informed multiscale wavelet decomposition agree well with human similarity judgments (e.g., Dailey, Cottrell, Padgett, & Adolphs, 2002; Kalocsai,

Zhao, & Biederman, 1998; Lyons & Akamatsu, 1998; Lyons, 2000). The extracted feature vectors contain information on oriented edges at multiple scales and form an efficient basis for object recognition (see, e.g., Rao & Ballard, 1995). To reduce the dimensionality of the feature vectors we used principal component analysis (PCA). PCA finds a linear transformation of n-dimensional vectors onto a space spanned by n orthogonal principal components. The components are ordered in terms of the variance they explain. The dimensionality of the extracted feature vectors is reduced by taking the projection onto the first p ($p < n$) principal components. The p-dimensional feature vectors so obtained reside in a similarity space where visual similarity translates to proximity of feature vectors.

Translating a two-dimensional image, using a multiscale wavelet decomposition followed by dimension reduction using PCA, is an often applied method in the domain of visual object recognition to model the first stages of processing of information in the human visual system (i.e., retina/LGN, V1/V2, V4; Dailey et al., 2002; Palmeri & Gauthier, 2004). In contrast, existing memory models lack such a preprocessing method and often make simplifying assumptions about item representations.

2.1.2. *Implementation of the perceptual preprocessing stage*

The features extracted from the face image consist of the responses of different derivatives-of-Gaussian filters. By applying a set of these filters at multiple scales and orientations, a representation of the face image area centered at the fixation location is obtained (Freeman & Adelson, 1991). To extract the feature vectors, the entire input image is transformed into a multiscale representation at four spatial scales. At every scale, the image is processed by four oriented filters in the orientations 0°, 45°, 90°, and 135°, using the steerable-pyramid transform (Freeman & Adelson, 1991). The choice for four orientations is assumed to be sufficient for the task at hand. The steerable-pyramid transform with four orientations is overcomplete, which means that it contains more information than necessary to reconstruct the original image. The human visual system appears to "filter" retinal images with many more orientations (Palmer, 1999).

The steerable-pyramid processing results in the 16 (4 scales × 4 orientations) filtered images shown on the lower left-hand corner of Fig. 1. The pixel values represent filter coefficients. Brighter pixels correspond to larger filter responses. From each of the 16 images a 7 × 7 window is selected, centered at a fixation location, and the 7 × 7 coefficients of the 16 images are placed in a vector. The sampling of 7 × 7 coefficients at each scale and orientation of the steerable pyramid is based on the following two considerations. First, it is assumed to reflect the way the human visual system samples the local neighborhood of fixated locations at multiple scales (Postma, Van den Herik, & Hudson, 1997). Second, it corresponds roughly to the estimates of the resolution of spatial attention (Nakayama, 1990). In addition to the filter coefficients, the pixel values of a 7 × 7 window of a low-resolution version of the image, centered at the fixation location, are appended to the vector. The 7 × 7 low-resolution subimage is included in the feature vector, because it contains absolute brightness information at the fixation location. This information is absent in the other 7 × 7 windows that contain the coefficients reflecting filter responses. Taken together, each fixation yields an 833-dimensional feature vector that contains information ranging from visual details (high-scale features) to coarse visual characteristics (low-scale features). It is worth noting that although information about the spatial locations and arrangement of fixations is not represented explicitly, it is represented implicitly by the low-scale features.

Fixation locations are randomly drawn from the contours of the face image. To detect the contours in an image we used an edge detector, which is based on the standard Canny edge detector with an adaptive threshold (i.e., the minimum and maximum intensity values) and $\sigma = 1$ (Canny, 1986). In Fig. 1 the contours of the input image are shown to the right of the input image. After feature-vector extraction, we reduce the dimensionality of the feature vectors extracted with the multiscale wavelet decomposition. To find the principal components that cover most of the variance of the 833-dimensional vectors, we applied PCA to a large set of feature vectors (i.e., about 100,000 feature vectors extracted from the images used in our simulations) in advance of our simulations. Then, in our simulations, we projected the feature vectors onto the first $p = 50$ of these principal components. It has been shown that approximately 50 components are sufficient to accurately represent faces (Calder, Burton, Miller, Young, & Akamatsu, 2001; Dailey et al., 2002; Hancock, Burton, & Bruce, 1996).

2.2. The memory stage

The memory stage of the NIM model is based on the GCM (Nosofsky, 1986, 1987). Although the GCM was initially introduced as an exemplar-similarity model for explaining categorization and identification of multidimensional perceptual items (e.g., Nosofsky, 1986, 1987), it has successfully been applied to old–new recognition tasks as well (e.g., Busey & Tunnicliff, 1999; Nosofsky & Zaki, 2003; Zaki & Nosofsky, 2001). GCM-based recognition models assume that individual exemplars are stored in memory and that recognition is based on the similarity of new inputs to previously stored inputs (see, e.g., Busey, 2001; Busey & Tunnicliff, 1999). In the following section, we briefly discuss the memory stage of the NIM model that is a recognition version of the GCM. Subsequently, we describe the implementation details.

2.2.1. The recognition version of the GCM

We distinguish two processes in the memory stage: the storage process and the recognition process.

2.2.1.1. The storage process. In the NIM model, exemplars of natural images are retained (i.e., "stored") in a similarity space. The storage process stores the preprocessed natural images. A preprocessed natural image is represented by a number of fixations, that is, low-dimensional feature vectors in a similarity space, each corresponding to the preprocessed image contents at the fixation location.

2.2.1.2. The recognition process. The recognition process determines the familiarity of an image by summing the similarities of the feature vectors of the image to previously stored feature vectors. The similarity of two feature vectors is defined as a decreasing function of their Euclidean distance in the similarity space. The similarity summed over all previously stored feature vectors serves as a basis for old–new recognition decisions.

Although, Nosofsky (1986) suggested that similarities are changed systematically by selective attention, we, for now, ignore the role of attention. This means that each dimension in the similarity space receives an equal weight.

2.2.2. Implementation of the memory stage

2.2.2.1. Implementation of the storage process. A preprocessed natural image is stored with storage strength *S,* which is defined as the number of feature vectors stored for an image (corresponding to the number of fixations). Fig. 2 is an illustration of the feature-vector representations of two natural images in a two-dimensional projection of a similarity space. The *x* axis and the *y* axis show the first and second dimensions—that is, principal components—respectively. (In the illustration, each image is represented by a cluster of 400 feature vectors—i.e., fixations, indicated by black dots [·] and gray crosses [×].) Note that this is an example of a representation based on 400 eye fixations. In our recognition simulations, however, we stored 10 fixations for each image. This corresponds to about 2 to 3 sec of viewing time in a recognition-memory experiment, as humans make approximately 3 to 5 eye fixations per second (see, e.g., Henderson, 2003; McSorley & Findlay, 2003).

2.2.2.2. Implementation of the recognition process. The familiarity of an image to be recognized is based on the average familiarity of its feature vectors. The familiarity of a feature vector is determined by the summed similarity to all the previously stored feature vectors. The GCM typically uses a Gaussian or an exponential decay function for relating stimulus similarity to distance in the similarity space. In a preliminary study, we obtained the best results with the step decay function. The step function is 1 when the Euclidean distance is smaller than or equal to a certain radius *r* and 0 otherwise. In other words, the step function simply calculates the number of previously stored feature vectors that reside within a hypersphere with radius *r* around a feature vector. Formally, the familiarity of a feature vector, *j,* is denoted as

$$fam_j = \sum_{i=1}^{T} s(d_{i,j}),\ s(d_{i,j}) = 1 \ \textit{if} \ \ d_{i,j} \leq r,\ 0 \ \textit{otherwise},$$

with T the total number of previously stored feature vectors and $d_{i,j}$ the Euclidean distance between feature vector *i* and *j.* Familiarity of an image to be recognized is defined as the average

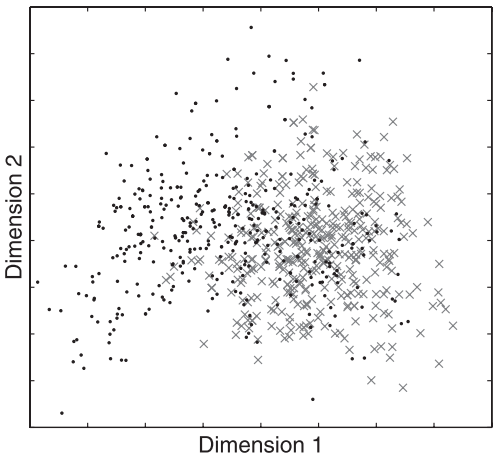

Fig. 2. Two-dimensional projection of the feature-vector representations of two natural images in a similarity space.

familiarity of its feature vectors. We use the logistic function to transform the familiarity values into a recognition probability between 0 and 1 (see, e.g., Busey, 2001; Busey & Tunnicliff, 1999). The logistic function is defined as

$$P(recognized \mid J) = \frac{1}{1 + \beta e^{-\theta fam_J}}$$

with β and θ as two free parameters and fam_J as the familiarity of face J (Busey & Tunnicliff, 1999).

3. Validation of the NIM model with individual natural stimuli

Because the key characteristic of the NIM model is that it transforms natural images into similarity-space representations, the psychological plausibility of the constructed similarity space is crucial for the model's behavior. To validate the plausibility of the similarity space, we tested the NIM model on a similarity-rating task. After that, we validated the NIM model's recognition predictions for individual natural stimuli on a recognition task. We compared the results of the NIM model with those obtained in corresponding behavioral experiments.

3.1. The similarity-rating task

We tested the psychological plausibility of the similarity space that the NIM model builds from natural input. To do so, we compared model similarity ratings with experimentally obtained human similarity ratings. Human similarity ratings were obtained in two studies: Busey and Arici (2005) and Busey (1998). In the rest of these subsections, we review the two behavioral similarity-rating experiments, describe the similarity-rating stimulations with the NIM model, and provide a discussion of the results.

3.1.1. The behavioral similarity-rating experiments

In both the experiments of Busey and Arici (2005) and Busey (1998), participants were repeatedly presented with two face images and were instructed to rate the similarity by assigning a number ranging from 1 (*most similar*) to 9 (*least similar*). This resulted in human similarity ratings for all possible pairs of faces. The sets of stimuli used in the experiments consisted of gray-scale images of bald male faces without glasses. Three examples of these faces are shown in Fig. 3.

In Busey and Arici (2005), 238 participants were tested on pairs of faces from a set of 60 faces (8 of the faces contained in the set were morphs, created by averaging 2 of the 52 other faces). Across participants, this resulted in a total of about 25 ratings for each of the 1,770 possible pairs of faces. In Busey (1998), 343 participants were tested on pairs of faces from a set of 104 faces (16 of which were morphs, created by averaging 2 of the 88 other faces), resulting in at least six ratings for each of the 5,356 possible pairs of faces. Within each experiment, the similarity ratings of individual participants were translated into z scores and averaged across participants.

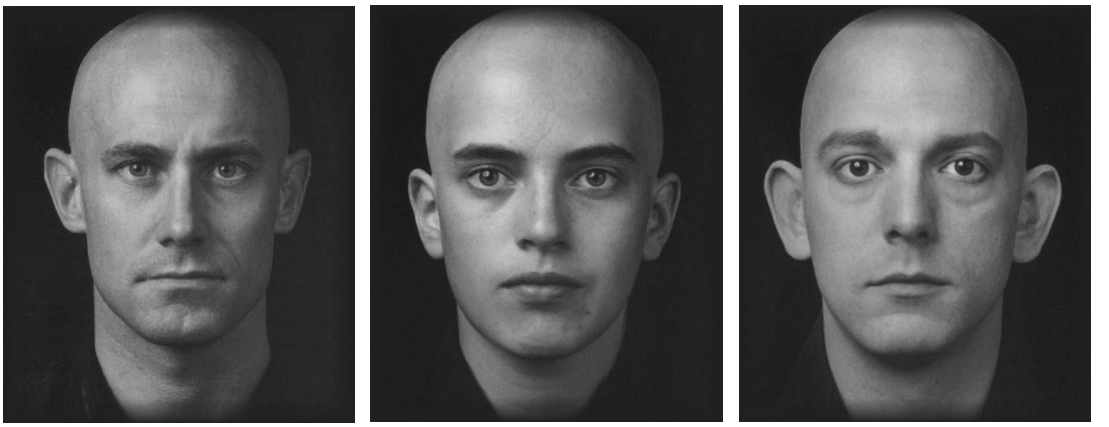

Fig. 3. Three examples of faces contained in the sets of faces used in the similarity-rating experiments.

3.1.2. Similarity-rating simulations with the NIM model

The NIM model bases a similarity rating for two faces on a matching of the similarity-space representations that result from preprocessing those faces. Because this task involves no memory processes, the memory stage is not employed in these simulations.

We present the NIM model with all possible face pairs from the sets of face images used in the experiments of Busey and Arici (2005) and Busey (1998). For the assessment of the similarity of two faces, we used a method similar to the assessment of familiarity as described previously. For each face of a pair (*A,B*), 100 feature vectors were extracted. Then, for each feature vector of face *A,* we determined the summed similarity to the feature vectors of face *B,* using the step function as defined for familiarity with radius parameter *r.* The average summed similarity defines the similarity value *s* for faces *A* and *B.* To compare the similarity values to the human similarity ratings, we linearly transformed the human similarity ratings (i.e., the average *z* scores of the human similarity ratings) to cover the range from 0 to 1 and logistically transformed the model similarity values into similarity ratings (*SR*), using the logistic function as defined previously. In our simulations, we varied the radius *r* from 2.0 to 6.0 to obtain the value for which correlations between human similarity ratings and model *SR* produced by the NIM model were optimal. Results were averaged across 1,000 simulation runs.

3.1.3. Similarity-rating simulation results

The highest correlations between model and human similarity ratings were found when the radius value was in the range $r = 4.0–5.4$. Fig. 4 presents the correlations between human and model *SR* as a function of the radius *r.*

As can be seen in Fig. 4, for the experiment of Busey and Arici (2005), the simulations with $r = 5.1$ produced the highest correlation of $R = .65$. For the experiment of Busey (1998), the simulations with $r = 4.9$ resulted in the highest correlation, which was $R = .70$. Fig. 5(a) shows the *SR* produced by the NIM model with radius $r = 5.1$ as a function of the human similarity ratings that were obtained experimentally by Busey and Arici (2005). Fig. 5(b) shows human and model similarity ratings (for $r = 4.9$) for the experiment of Busey (1998).

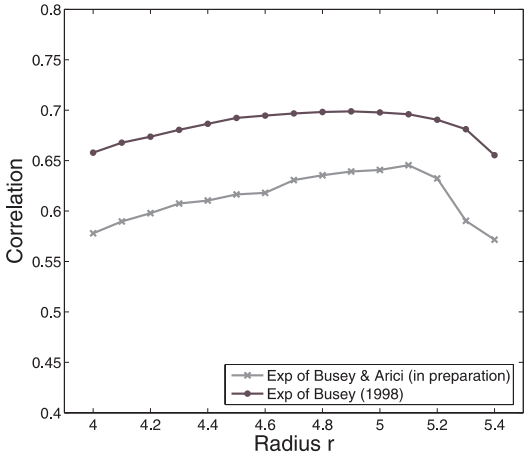

Fig. 4. Correlations between the NIM model's similarity ratings (*SR*) and experimentally obtained human similarity ratings as a function of the radius *r*.

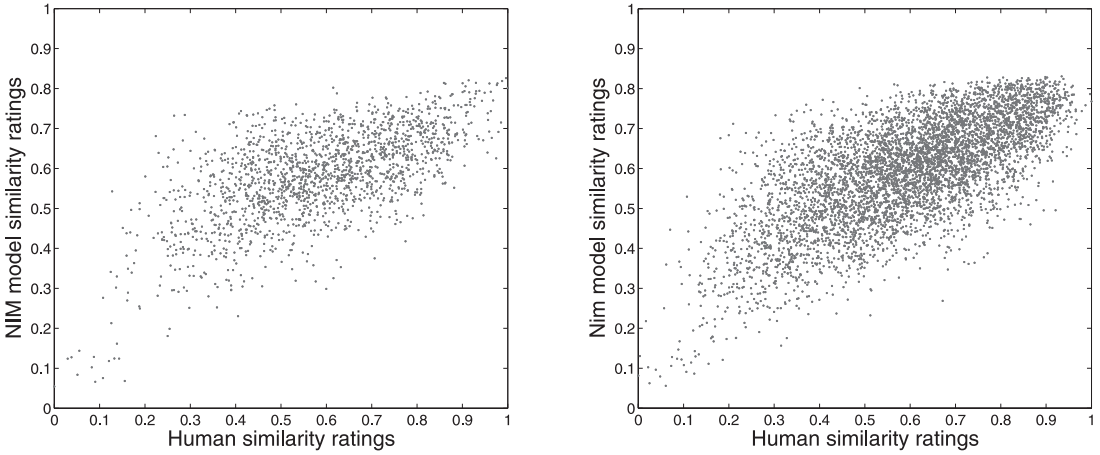

Fig. 5. The NIM model's similarity ratings (SR) as a function of the experimentally obtained human similarity ratings (linearly transformed to the range 0 to 1). (a) The experiment of Busey and Arici (2005), $R = .65$. (b) The experiment of Busey (1998), $R = .70$.

3.1.4. Discussion of the similarity-rating modeling results

Here we briefly discuss four points that offer some perspective on the correlations obtained in our experiments: (a) the limited consistency of similarity judgments across participants, (b) the logistic transformation of similarity values into similarity ratings, (c) the selection of eye-fixation locations, and (d) the contribution from the different spatial scales. Then, we relate our results to the results of various other studies that have also compared the representation

spaces that resulted from applying an image preprocessing scheme to the outcomes of psychological experiments.

The first point concerns the limited consistency of the human similarity judgments. An analysis of the human similarity ratings obtained in the experiment of Busey and Arici (2005) showed that there was considerable variation between different participant's similarity ratings. When similarity ratings of the participants were divided into five groups (each consisting of some 48 participants), correlations between the different group averages ranged from 0.72 to 0.75. This demonstrates the limited consistency of the similarity judgments across participants. It is, therefore, highly unlikely that correlations between similarity ratings produced by our model and experimental similarity ratings will exceed the value of 0.75. Given this consideration, the similarity-space representations of the NIM model can be said to correlate reasonably well with the experimentally obtained similarity ratings.

The second point is the use of a nonlinear function to map similarity values onto similarity ratings. Our motivation for using the nonlinear logistic function is that it constrains similarity ratings to the unit interval (cf. Busey & Tunnicliff, 1999). Because the logistic function produces a constrained range of outputs, we applied it to translate the NIM model's similarity values into similarity ratings. However, a linear transform of the similarity values produced comparable correlations that were 0.64 and 0.70 for the experiments of Busey and Arici (2005) and (Busey, 1998), respectively.

The third point is about the role of eye-fixation locations. The correlations obtained can be explained partly by the selection of eye-fixation locations. In a separate study, we selected eye-fixation locations randomly along the image (as opposed to selection along the contours in the image). Using this procedure, correlations between model and human similarity ratings are significantly lower than when fixations are selected along the contours in the image. The selection of eye-fixation locations plays an important role in human vision (e.g., Rajashekar, Cormack, & Bovik, 2002; also see General discussion section).

Finally, the fourth point is about the contribution from the various spatial scales. To test this contribution, we ran similarity-rating simulations that selectively ignored the features at one of the four spatial scales. The results demonstrate that correlations between model and human similarity judgments decrease more significantly when low-scale visual features (i.e., coarse visual information) are removed than when high-scale visual features (i.e., visual details) are removed. It is very well possible that coarse visual information plays a more important role than detailed visual information when judging the similarity of two faces.

Two other studies have compared representation spaces with the human similarity ratings obtained in the behavioral studies described in this section. Dailey, Cottrell, and Busey (1999) performed a comparison with the 104 face images used in the experiment of Busey (1998), and Steyvers and Busey (2000) with the 60 face images used in the experiment of Busey and Arici (2005). We briefly discuss both studies. The study of Dailey et al. (1999) examined the ability of three different memory models operating on three types of representation spaces to account for the experimental face-recognition results of Busey and Tunnicliff (1999). Two of these three types of representation spaces were generated using PCA and Gabor filtering followed by PCA (similar to the preprocessing in the EMPATH model of Dailey et al., 2002; see the General discussion section for a detailed description of the differences between the NIM

model and the EMPATH model). They compared the resulting representations with the distances between pairs of faces in multidimensional scaling (MDS) space for the 104 face images. They found a correlation of .388 for the principal component representation space and a correlation of .517 for the Gabor-filter representation space, which are considerably smaller than the correlation of 0.70 obtained with the NIM model.

Higher correlations were obtained by Steyvers and Busey (2000). In their feature-mapping model, Steyvers and Busey related the features extracted with various preprocessing mechanisms to the dimensions of an MDS solution based on human similarity judgments for the 60 face images (Busey & Arici, 2005). They investigated three different preprocessing mechanisms: PCA, Gabor-filter-based preprocessing, and geometric information extraction (i.e., describing the face by a set of distances between landmark points on the face). The feature-mapping model differs from the NIM model in an important respect: The preprocessed images are fed into an optimization network to enhance the fit with behavioral data. The optimization network learned the mapping between the input features and the psychological dimensions of the MDS solution of the human similarity ratings (Steyvers & Busey, 2000). Considering that Steyvers and Busey specifically optimized their representations with respect to the human similarity ratings, it is not surprising that they found respectable correlations ranging from 0.45 to 0.86. In contrast to their approach, the NIM model changes only three parameters, whereas Steyvers and Busey adapted at least 40 weights to optimize the fit with the human data. The study of Steyvers and Busey provides important clues about how the dimensions of feature vectors extracted with a certain preprocessing mechanism should be weighed in a psychologically plausible way. Future versions of the NIM model should similarly address the weighing of feature-vector dimensions in a more psychologically plausible way.

Several other studies have compared representation spaces resulting from various image preprocessing schemes to the outcomes of psychological experiments. For instance, Calder et al. (2001) and Dailey et al. (2002) did this for coding facial expressions and Hancock, Bruce, and Burton (1998), Kalocsai et al. (1998), and Lyons (2000) did this for face identity. Hancock et al. (1998) compared human similarity judgments with similarity-space representations based on a PCA or on a graph-matching system. They found small correlations of about 0.20 and argued it to be caused by the noisiness of the human data. Kalocsai et al. used a same–different judgment task to compare the performances of a preprocessing scheme based on Gabor-filtering and a global template-matching classifier with human data. In the experiment, participants had to judge whether two sequentially presented images were of the same individual. Kalocsai et al. found correlations up to 0.91. However, these were based on a small sample of 16 judgments. Moreover, the similarity judgments were obtained using a much simpler binary classification task. Lyons found a correlation of 0.71 between human and model-based face similarity ratings, which is slightly higher than the correlations we found. However, he employed a very small set of 10 facial images and did not apply PCA to reduce the dimensionality of the representation space.

3.2. The recognition task

To validate the recognition predictions of the NIM model for individual natural stimuli, we compared the recognition rates produced by the NIM model with those that were obtained in

two behavioral experiments. The first behavioral recognition experiment is the face-recognition experiment of Busey and Arici (2005), employing 60 faces and 238 subjects. The second is the face-recognition experiment of Busey and Tunnicliff (1999), employing 104 faces and 180 subjects. The face images used in these experiments are identical to those used in the similarity-rating experiments.

In the following subsections, we describe the behavioral face-recognition experiments, present the simulation results obtained with the NIM model, and provide a discussion.

3.2.1. Behavioral recognition experiments

In their recognition experiments, Busey and Arici (2005) and Busey and Tunnicliff (1999) assessed the recognition rates (i.e., hit rates and false-alarm rates) for different types of faces. The sets of 60 and 104 faces employed in the experiments contained two types of faces: (a) normal faces and (b) morph faces. Each morph face was the average of two normal so-called parent faces.

In the face-recognition experiment of Busey and Arici (unpublished results), the set of 60 faces was subdivided into two sets of 30 faces such that each set contained 26 normal faces, 8 of which were defined as parent faces, and 4 as morph faces. When a morph was in one of the two sets, its parents were in the other. Half of the subjects were presented with a study list with the faces from Set 1 and tested for old–new recognition for the faces from both Set 1 (i.e., targets) and Set 2 (i.e., lures). For the other half of the subjects this was reversed. This allowed hit and false-alarm rates to be obtained for each of the 60 faces.

In the face-recognition experiment of Busey and Tunnicliff (1999), participants were presented with a study list with 68 normal faces, 32 of which were defined as parent faces. Then, old–new recognition was tested for the 68 normal faces from the study list (i.e., 36 normal targets and 32 parent targets) along with 20 new normal faces (i.e., normal lures) and 16 new morph faces (i.e., morph lures). The morph lures were the average of two parent targets that were either dissimilar or similar to each other. Dissimilar and similar morph lures resulted from dissimilar and similar morph parents, respectively.

3.2.2. Recognition simulations with the NIM model

The NIM model simulations use study and test lists identical to those used in the behavioral experiments. In the simulations of Busey and Arici's (2005) face-recognition experiment, the NIM model was presented with the 26 normal faces, 8 of which were parent faces, and 4 morph faces from Set 1 in half of the simulations (henceforth referred to as data set A) and from Set 2 in the other half (data set B). Then the model was tested for old–new recognition of the presented faces (i.e., targets) along with the 26 normal faces, containing 8 parent faces, and 4 morph faces from the other set (i.e., lures).

In the simulations of the face-recognition experiment of Busey and Tunnicliff (1999), the model was presented with the 68 normal faces, 32 of which were parent faces from the study list of the behavioral study. Then recognition was tested for these faces along with the 20 new normal faces (i.e., normal lures) and the 16 new morph faces (i.e., morph lures) from the test list of the behavioral experiment.

In the behavioral experiments, the images were presented for 1,500 msec, followed by a 2-sec delay. In our simulations, the number of fixations selected and stored for each face was

set to 10, which corresponds to approximately 2 sec of viewing time (see, e.g., Henderson, 2003; McSorley & Findlay, 2003). For recognition, the NIM model calculated the familiarity of each target and lure on the basis of 100 fixations (as described previously). Familiarity values were transformed into recognition probabilities using the logistic transform as defined previously. In the simulations the radius *r* was varied from 2.0 to 6.0 to determine the *r* value that produces the smallest root mean square errors (RMSEs) and the largest correlation values between recognition rates produced by the NIM model and experimentally obtained human recognition rates (i.e., hit rates and false-alarm rates).

3.2.3. Recognition simulation results

The smallest *RMSE*s and the largest correlation values between the NIM model's recognition rates and human recognition rates were obtained with a radius in the range of $r = 2.6–3.6$. Fig. 6(a) presents the *RMSE*s and Fig. 6(b) the correlations between model and experimentally obtained recognition rates as a function of the radius *r*.

For the simulations of the face-recognition experiment of Busey and Arici (2005), the smallest *RMSE*s between experimentally obtained recognition rates and those produced by the NIM model were obtained for $r = 3.0$ ($RMSE = 0.155$, for data set A, and $RMSE = 0.160$, for data set B). The correlations for $r = 3.0$ were $R = .82$ and $R = .70$ for data sets A and B, respectively. Using this radius, Fig. 7(a) and 7(b) show the model recognition rates as a function of experimentally obtained human recognition rates for data sets A and B, respectively.

For the simulations of the face-recognition experiment of Busey and Tunnicliff (1999), the smallest *RMSE* and the highest correlation between experimentally obtained recognition rates and those produced by the NIM model also resulted in a radius $r = 3.0$: $RMSE = 0.151$ and $R = .66$.

Fig. 8(a) shows the recognition rates produced by the NIM model for $r = 3.0$ as a function of the human recognition rates (Busey & Tunnicliff, 1999).

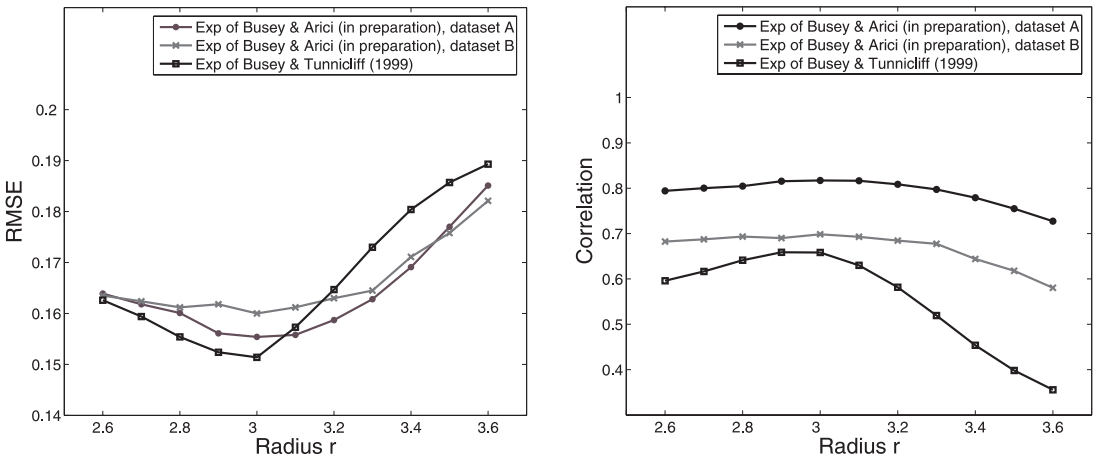

Fig. 6. A comparison of model recognition rates and human recognition rates obtained in Busey and Arici (2005). (a) The *RMSE* between model and human recognition rates as a function of the radius *r*. (b) The correlation between model and human recognition rates as a function of the radius *r*.

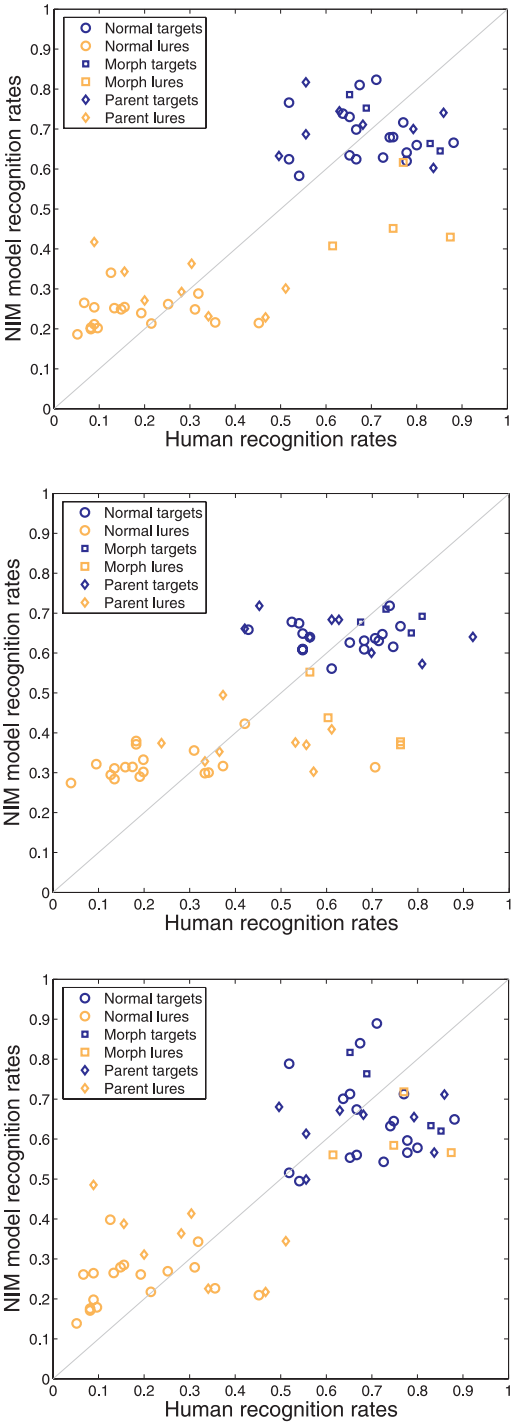

Fig. 7. The NIM model's recognition rates as a function of the human recognition rates (Busey & Arici, 2005). (a) Data set A, with $r = 3.0$, $RMSE = 0.155$, $R = .82$; (b) data set B, with $r = 3.0$, $RMSE = 0.160$, $R = .70$; (c) data set A, with $r = 3.3$, $RMSE = 0.162$, $R = .80$.

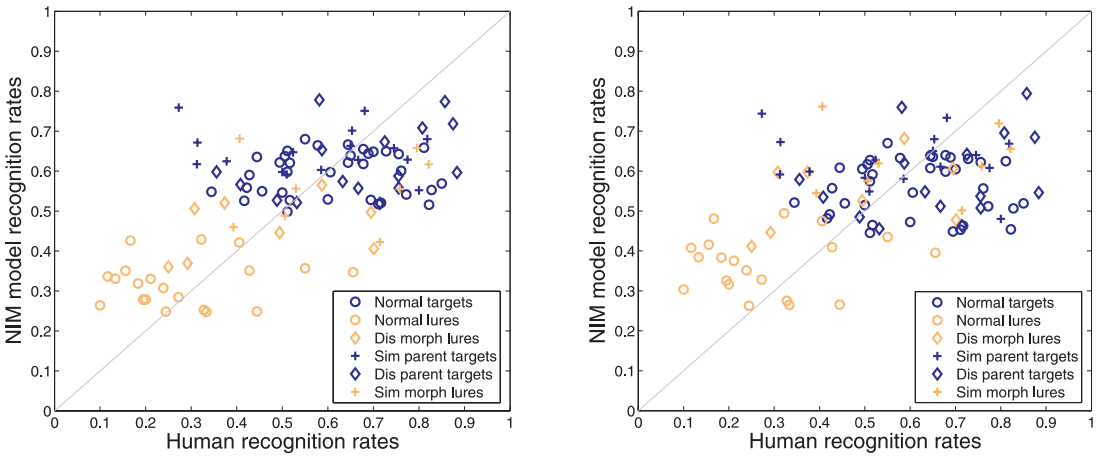

Fig. 8. The NIM model's recognition rates as a function of the human recognition rates (Busey & Tunnicliff, 1999). (a) With $r = 3.0$, $RMSE = 0.151$, $R = .66$; (b) with $r = 3.2$, $RMSE = 0.164$, $R = .58$.

3.2.4. *Discussion of the recognition modeling results*

The recognition rates produced by the NIM model agree quite well with experimentally obtained human recognition rates. In the following, the main differences and agreements between the recognition results of the NIM model and the experimentally obtained human recognition results are demonstrated. We start by identifying three main behavioral effects that are successfully replicated by the NIM model. Then, we briefly touch on the ability of the NIM model to explain the within-class variability of human recognition rates. Subsequently, we discuss to what extent our results depend on the number and locations of fixations, and on the different spatial scales. Finally, we compare our modeling results with those obtained in other face-recognition modeling studies (Busey & Tunnicliff, 1999; Dailey et al., 1999).

The first effect found in the experimental data is that morph lures are falsely recognized more often than normal lures. This effect was successfully produced by the NIM model for simulations of both the experiments (Busey & Arici, 2005; Busey & Tunnicliff, 1999).

The second effect is that in the experiment of Busey and Tunnicliff (1999), which employed similar and dissimilar morph lures, false-alarm rates for the similar morph lures were higher than those for the dissimilar morph lures. This effect was also produced successfully by the NIM model simulations of Busey and Tunnicliff's experiment.

The third effect concerns the false-alarm rates for lures compared to the hit rates for their parents. In the behavioral experiments the false-alarm rate for the morph lures approached the hit rates for their parents. In the experiment of Busey and Tunnicliff (1999) the false-alarm rates even marginally exceeded the hit rates for the similar lures (and not for the dissimilar lures). This effect is called the *morph-inversion effect* (Dailey et al., 1999). As can be seen in Figs. 7(a), 7(b), and 8(a), the NIM model produces substantially smaller false-alarm rates for the morph lures compared to the hit rates for their parents for $r = 3.0$. By increasing the radius r of the NIM model, a better agreement with experimental findings for the morphs can be obtained. This is at the expense, however, of a decrease in overall correlation and an increase in

the *RMSE*s between experimental and model recognition rates. Fig. 7(c) shows the results with parameter settings $r = 3.3$ for simulations of the experiment by Busey and Arici (2005; Sets 1 and 2 are targets and lures, respectively), and Fig. 8(b) shows the results for simulations of the experiment of Busey and Tunnicliff with radius $r = 3.2$. Compared to the results in Figs. 7(a) and 8(a), the false-alarm rates for the morph lures have increased, whereas the hit rates for their parents have decreased. Obviously, the increase in r selectively benefits the recognition rates for the different types of faces. With the increased radius, the NIM model findings agree with experimental findings of Busey and Tunnicliff; a morph-inversion effect is produced for the similar morph lures, but not for the dissimilar morph lures. This is a direct consequence of the face similarity structure reflected in the NIM model's similarity space. With a larger radius r, more stored fixations of the parent faces contribute to the familiarity of the morph lures, particularly when the target parents are highly similar to the morph lures (i.e., their feature-vector representations are close together in the similarity space). Because, the feature vectors of lures that are dissimilar to the stored targets are farther apart in the similarity space, the familiarity of these faces is only marginally affected by an increase in r.

Although quite high overall correlations between human and NIM model recognition rates are obtained, there is an important discrepancy between the human recognition rates and those produced by the NIM model. The discrepancy concerns the within-class variability of recognition rates (for example, within the class of normal targets). The human recognition rates show noticeably larger within-class variability than the NIM model recognition rates. Busey and Tunnicliff (1999) reported the same discrepancy between the human data and their SimSample model predictions based on an MDS space. They argued that the discrepancy is likely to be due to memory characteristics that operate independently from the overall similarity structure of the faces (e.g., a small, striking facial feature, such as a birthmark, may make the face highly memorable, but may be insignificant for the overall similarity of faces). Similarly, a different dependency on the similarity structure of faces might explain why the radius r for which the correspondence between the model and the human results is optimal, differs for the similarity-rating task and for the recognition task.

The correlations between human recognition rates and those produced by the NIM model appear to be relatively independent of the number of fixations that are selected and stored (i.e., the storage strength). In our original simulations we used a storage strength of 10 fixations, because this approximately corresponds to the number of fixations per face in a face-recognition experiment. However, to test to what extent the results depend on the storage strength, we ran simulations with various storage strengths. For both the experiments of Busey and Arici (2005) and Busey and Tunnicliff (1999), the different storage strengths (varying from 3 to 30 fixations) gave results similar to those obtained from the simulations with 10 fixations. Although, the correspondence between the NIM models and human recognition performance is unaffected by the number of stored fixations, the location of the fixations seems to play an important role. Randomly selecting fixations across the image rather than along the contours significantly reduces correlations between model and human recognition performance. The reduction mainly results from a substantial decrease in the ability to account for the variation of recognition rates across the different types of lures and across the different types of targets. As for the similarity-rating results, we analyzed the contribution from visual information at the different spatial scales. The decrease in correlations between model and human recognition

rates that results from removing visual information from one scale was approximately equal for the different spatial scales. In contrast, for the similarity-rating task, information from the lower spatial scales seemed to play a more important role than fine visual details. Apparently, the amount of visual detail used to make recognition decisions differs from the amount of detail used to make similarity judgments. This complies with Busey and Tunnicliff's suggestion that striking visual details can play an important role for memory, yet be unimportant for judging similarity.

The face-recognition results of Busey and Tunnicliff (1999) have been modeled previously by them and by Dailey et al. (1999). To explain the experimental recognition data, Busey and Tunnicliff tested several models among which were a recognition version of the GCM model, their SimSample model, and the SimSample model extended with two different versions of a prototype mechanism. They obtained *RMSE*s of 0.1800, 0.1462, 0.1441, and 0.1411 for the four models, respectively. The *RMSE*s are somewhat smaller than the *RMSE* of 0.151 obtained with the NIM model. This is likely due to the use of a weighed MDS space based on human similarity ratings instead of using the face images as input. In contrast, Dailey et al. (1999) used the face images as input for a PCA and a Gabor-filtering preprocessing mechanism. The resulting representations were used to test the ability of different memory models to account for the recognition data of Busey and Tunnicliff. Using the Gabor-filtering preprocessing method, they obtained *RMSE*s of 0.1624 with a version of the GCM.

4. General discussion

The main difference between the NIM model and existing memory models is that it encompasses a biologically informed perceptual front end that operates on natural images. In this section, we discuss the perceptual front end, possible extensions for the NIM model, and present a concluding remark.

4.1. A biologically informed front end to operate on natural input

The perceptual preprocessing in the NIM model applies a transformation that yields a perceptual similarity structure of natural images. This section discusses how the NIM model relates to computational memory models that lack a perceptual preprocessing method, compares the NIM model's preprocessing method with the widely applied image-processing method of PCA and with the EMPATH model of Dailey et al. (2002), and discusses how our perceptually based similarity space differs from a conceptually based similarity space.

4.1.1. The benefit of a perceptual front end

So far, existing memory models have been tested with artificial data (e.g., the REM model; Shiffrin & Steyvers, 1997), with predefined similarity spaces (e.g., the SimSample model, Busey, 2001; the GCM, Nosofsky, 1987; the REM model, Steyvers, 2000), and with synthesized texture images (the NEMO model, Kahana & Sekuler, 2002). For conceptually based stimuli, many methods for defining the similarity space are based on word co-occurrence counts in texts (latent semantic analysis; e.g., Derweester, Dumais, Furnas, Landauer, &

Harshman, 1990; Landauer & Dumais, 1997; hyperspace analog to language; e.g., Burgess, Livesay, & Lund, 1997), on experimentally obtained free association data (word association spaces; e.g., Steyvers et al., 2004), or on experimentally obtained similarity ratings (MDS; Caramazza, Hersch, & Torgerson, 1976). For perceptually based stimuli, such as faces, the latter approach has also been used often (e.g., Busey, 2001). Because the resulting perceptual similarity spaces accurately reflect similarities as perceived by humans, this approach leads to useful representational models of memory. However, because representations are not derived directly from the individual input patterns, these models fall short in constructing a representation that is grounded in the real world. Because the preprocessing stage in the NIM model can be conceived as an image-processing front end, it can be applied to other models of memory to realize grounded representations. Therefore, the NIM model's front end complements rather than replaces existing computational memory models such as the REM model. Because these memory models base recognition decisions on the similarity to stored exemplars, we expect that they will also produce the results that were described in this article, given that they accurately represent the similarity structure of the face images used in this study. The strength of the NIM model lies in the fact that it is able to produce the recognition results based on the use of natural images. The results presented in this article reveal that the NIM model's recognition rates for individual face images correlate well with those obtained in behavioral experiments. Therefore, the NIM model can be used to predict recognition judgments for novel stimuli. We consider this a clear benefit of a perceptual front end.

4.1.2. Image processing

The approaches to modeling the cognitive aspects of recognition have largely evolved independently of the image-processing approaches to recognition. The latter are mainly concerned with how visual input can be mapped onto a certain representation despite variations in viewpoint and lighting conditions. The NIM model introduces advanced image preprocessing in the cognitive modeling of memory. The advantages of combining image-processing techniques with cognitive approaches have been emphasized by several researchers (Burton, Bruce, & Hancock, 1999; Calder et al., 2001; Dailey et al., 2002; Edelman, 1995; Steyvers & Busey, 2000). The method of PCA is widely applied in the domain of image processing. In the following, we consider how our preprocessing method relates to PCA. Then, we discuss how the NIM model's preprocessing stage relates to that of the EMPATH model of Dailey et al. (2002).

PCA applied to the pixel values of natural images yields a sparse representation of the images (Hancock, Baddeley, & Smith, 1992). A number of face-recognition models (e.g., Burton et al., 1999; O'Toole, Abdi, Deffenbacher, & Valentin, 1993; Turk & Pentland, 1991) apply PCA to the entire (shape-standardized) image to obtain so-called eigenfaces, the principal components that account for most of the variance in a number of face images. Like the similarity space in the NIM model, the similarity space spanned by the eigenfaces forms the psychological abstract notion of a similarity space for faces that is generally assumed to underlie face memory (O'Toole, Wenger, & Townsend, 2001; Valentine, 1991). Models based on principal components have been shown to successfully perform different recognition tasks, such as, old–new recognition memory, identification, and recognition of facial expressions (e.g., Burton et al., 1999; Calder et al., 2001; O'Toole et al., 1993). Although PCA applied to the entire face image extracts important visual features from a set of face images, the nature of the fea-

tures depends on the specific expressions or shapes of the faces. Therefore, the use of a different set of images requires recomputation of the principal components so that they fit the new set. To obtain the general features of natural images, a PCA should be applied to an extensive collection of natural images containing a wide variety of objects and scenes instead of a limited set of training images. Hancock et al. (1992) found the principal components that resulted from performing PCA on a large number of natural images to approximate derivatives of two-dimensional Gaussian functions (or Gabor functions). Such functions, which are used for feature extraction in the NIM model, form an appropriate basis for building a similarity-space representation from natural images that is independent of the specific set of images used. Therefore, we prefer their use over standard PCA applied to the raw or shape-standardized image. After feature vectors have been extracted using Gaussian derivatives, we still apply a PCA to the extracted feature vectors to reduce computational demands.

Our preprocessing stage resembles that of the EMPATH model developed by Dailey et al. (2002): multiscale wavelet filtering followed by PCA. The EMPATH model relies on the preprocessing to explain psychological findings on the perception of facial expressions. Moreover, as discussed previously, the model has been used as a preprocessing front end to different memory models in an attempt to account for the experimental recognition data of Busey and Tunnicliff (1999; Dailey et al., 1999). By showing that their EMPATH model simulates a variety of psychological results on categorization, similarity, discrimination, and recognition difficulty related to facial expression perception, Dailey et al. (2002) validated the similarity structure extracted from the input by the preprocessing method. Although the image-processing methods applied by the NIM model and the EMPATH model are similar, they differ in two ways. The first difference is a minor one. For feature extraction, Dailey et al. (2002) used Gabor wavelets at different scales and orientations. In contrast, we use the steerable pyramid (a set of Gaussian derivatives) at different scales and orientations. Although, Dailey et al. (2002) used a slightly different set of filters, both feature-extraction methods are multiscale wavelet decompositions that contain information about oriented edges at different scales and orientations. The second and main difference between the NIM and EMPATH models concerns the way in which image locations are selected for feature extraction. In the EMPATH model, images are filtered at 29×35 grid points, evenly distributed over the image (in Dailey et al., 1999, only 64 grid points were used). In contrast, the NIM model selects a few eye-fixation locations (in our simulations we used 10 fixations) randomly along the contours in the image and extracts the filter responses from a small image area surrounding each fixation. This corresponds to the way human subjects attend different parts of a visual scene by means of eye fixations. The NIM model's mechanism of selectively fixating different parts of the image provides a natural way of dealing with overt spatial attention. Moreover, because the number of stored fixations corresponds to viewing time, the timing of stimuli can be modeled. Simulations that manipulate the viewing time of the NIM model show that the model's recognition performance increases with the number of stored fixations (Lacroix, Murre, Postma, & Van den Herik, 2004).

4.1.3. Perceptually based versus conceptually based similarity spaces

The preprocessing method employed by the NIM model is specifically suitable to derive a perceptually based similarity space. There is a long-standing debate on the question as to what extent similarity-related processes, such as category formation and inductive generalization,

are based on perceptual similarity or on conceptual similarity. Several researchers have discussed the powerful nature of perceptually based representations (e.g., Goldstone & Barsalou, 1998; Goldstone, Steyvers, Spencer-Smith, & Kersten, 2000) especially for children (Hayes & Heit, 2004). Although we realize that perceptually based similarity alone can hardly constitute a plausible basis for explaining all similarity-related memory phenomena, it plays an important role in a wide range of situations. In the NIM model, the similarity of novel human faces is assumed to be based on the perceptual features of the faces. Clearly, the assumption would not be warranted for words where the similarity is known to be based mainly on conceptual and much less on perceptual characteristics. In our view, words (concepts) and images (objects) are preprocessed separately to yield psychologically plausible similarity spaces for concepts and images, respectively. When their similarities are represented properly, both spaces will exhibit the effects examined in our article when operated on by the memory stage. The effects follow from the similarity-space representations, rather than from the modality that gave rise to these representations.

4.2. Possible extensions

In its present form, the NIM model is simple and straightforward. Broadening its modeling capabilities requires the model to be extended. In the following we discuss two extensions that endow the NIM model with a feature-based and spatial attentional mechanism.

4.2.1. Feature-based attention

The first possible extension is the task-dependent adaptation of the distances in the NIM model's similarity space by stretching or shrinking the axes. It has been shown that humans attend to different features depending on task-related factors (e.g., Goldstone & Steyvers, 2001; Halberstadt, Goldstone, & Levine, 2003; Nosofsky, 1987) or on individual differences in perceptual or attentional processes (e.g., Halberstadt et al., 2003; Viken, Treat, Nosofsky,McFall, & Palmeri, 2002). In the NIM model, a feature-based attentional mechanism might adjust the similarity space by assigning those dimensions of the similarity space that are of relevance a higher weight than those that are irrelevant. The weighing of dimensions can be based on knowledge obtained in experimental studies that examine the role of information at different spatial scales for performing a certain task (see, e.g., Goffaux, Hault, Michel, Vuong, & Rossion, 2005). This knowledge can be incorporated into the preprocessing stage such that different parts of the vector (that contain information on different scales) are weighed differently. Also the weighing of the different feature-vector parts can be derived from human similarity-rating data (Steyvers & Busey, 2000) and recognition data. This might result in more psychologically plausible similarity-space representations.

4.2.2. Spatial attention

The second possible extension of the NIM model is the spatial selection of interesting regions for visual information extraction. Although the weighing of dimensions (as described previously) provides an appropriate way to enhance the psychological plausibility of the similarity space, the NIM model has a natural way of dealing with spatially based attention by fixating different image regions (i.e., overt attention). Currently, the NIM model selects an eye

fixation along the contours in the image, independent of its current or past states. In contrast, human visual selection is context-dependent (e.g., Rajashekar et al., 2002). In the dynamic process of actively scanning the visual scene, eye fixations are guided by covert attention that includes bottom-up and top-down processes (Henderson, 2003; Karn & HayHoe, 2000; Oliva, Torralba, Castelhano, & Henderson, 2003). The NIM model might be equipped with a mechanism that selects image regions on the basis of conspicuous features or stored knowledge.

4.3. Concluding remark

The NIM model builds a similarity space from natural input and incorporates a recognition version of the GCM. In this article, the NIM model is validated with individual natural stimuli. The model is tested on a similarity-rating task and a face-recognition task using the same face images as were used in experimental studies. From our results we conclude that the NIM model can simulate similarity ratings and recognition performances for individual natural stimuli quite reliably. Complementing a memory model with a perceptual front end allows for automatic predictions of recognition memorability for individual novel stimuli.

Notes

1. In principle, the steerable pyramid can be used to detect the contours.
2. Also referred to in Steyvers and Busey (2000).
3. One point should be noted about the third effect. Although a morph-inversion effect was obtained for the similar morph lures (but not for the dissimilar morph lures) in the behavioral experiment of Busey and Tunnicliff (1999), the effect was not statistically significant ($\alpha > 0.05$), and the average recognition rates for the similar morph lures only marginally exceeded the average recognition rates for their parents.

Acknowledgments

Dr. Busey of Indiana University Bloomington, is gratefully acknowledged for providing us with his data set of facial images and helpful comments. Moreover, we wish to thank the reviewers for their helpful comments on earlier versions of this article. The research project (Project Number 051.02.2002) is supported in the framework of the NWO Cognition Program with financial aid from the Netherlands Organization for Scientific Research (NWO).

References

Arkadev, A. G., & Braverman, E. M. (1966). *Computers and pattern recognition.* Washington, DC: Thompson.
Barlow, H. B. (1989). Unsupervised learning. *Neural Computation, 1,* 295–311.
Bellman, R. (1961). *Adaptive control processes: A guided tour.* Princeton, NJ: Princeton University Press.
Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, England: Oxford University Press.

Burgess, C., Livesay, K., & Lund, K. (1997) Explorations in context space: Words, sentences, discourse. *Discourse Processes, 25,* 211–257.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science, 23,* 1–31.

Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science, 9,* 476–482.

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. In M. Wenger & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 147–192). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Busey, T. A., & Arici, A. (2005). Dissociations of accuracy and confidence reveal the role of individual items and distinctiveness in recognition memory and metacognition. Manuscript in preparation.

Busey, T. A., & Tunnicliff, J. (1999). Accounts of blending, typicality and distinctiveness in face recognition. *Journal of Experimental Psychology: Learning Memory and Cognition, 25,* 1210–1235.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research, 41,* 1179–1208.

Canny, J. (1986). A computational approach to edge detection. IEEE *Transactions on Pattern Analysis and Machine Intelligence, PAMI-8,* 679–698.

Caramazza, A., Hersch, H., & Torgerson, W. S. (1976). Subjective structures and operations in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 15,* 103–117.

Dailey, M. N., Cottrell, G. W., & Busey, T. A. (1999). Facial memory is kernel density estimation (almost). In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 24–30). Cambridge, MA: MIT Press.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14,* 1158–1173.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478.

Derweester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41,* 391–407.

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and Machines, 5,* 45–68.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences, 21,* 449–498.

Edelman, S., & Intrator, N. (1997). Learning as extraction of low-dimensional representations. In R. Goldstone, D. Medin, & P. Schyns (Eds.), *Mechanisms of perceptual learning* (Vol. 36, pp. 353–380). San Diego, CA: Academic.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration and charm. *Psychological Review, 92,* 1–38.

Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,* 891–906.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34,* 77–86.

Goldstone, R., Steyvers, M., Spencer-Smith, J., & Kersten, A. (2000). Interactions between perceptual and conceptual learning. In E. Diettrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 191–228). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65,* 231–262.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130,* 116–139.

Halberstadt, J., Goldstone, R. L., & Levine, G. M. (2003). Featural processing in face preferences. *Journal of Experimental Social Psychology, 39,* 270–278.

Hancock, P. J. B., Baddeley, R. J., & Smith, L. S. (1992). Principal components of natural images. *Network, 3,* 61–70.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research, 38,* 2277–2288.

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

Hayes, B. K., & Heit, E. (2004). Why learning and development can lead to poorer recognition memory. *Trends in Cognitive Sciences, 8,* 337–339.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Science, 7,* 498–504.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hubel, D. H. (1988). *Eye, brain, and vision.* New York: Freeman.

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

Kalocsai, P., Zhao, W., & Biederman, I. (1998). Face similarity space as perceived by humans and artificial systems. In *Proceedings, Third International Conference on Automatic Face and Gesture Recognition* (pp. 177–180). Los Alamitos, CA: IEEE Computer Society.

Karn, K. S., & HayHoe, M. M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition, 7,* 673–703.

Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., & Van den Herik, H. J. (2004). The natural input memory model. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society (CogSci 2004)* (pp. 773–778). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lee, T. S. (1998). Image representation using 2d Gabor wavelets. *IEEE Transactions of Pattern Analysis and Machine Intelligence, 18,* 959–971.

Lyons, M., & Akamatsu, S. (1998). Coding facial expressions with Gabor wavelets. In *Proceedings, Third International Conference on Automatic Face and Gesture Recognition* (pp. 200–205). Los Alamitos, CA: IEEE Computer Society.

Lyons, M. J. (2000). A linked aggregate code for processing faces. *Pragmatics and Cognition, 8,* 63–81.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760.

McSorley, E., & Findlay, J. M. (2003). Saccade target selection in visual search: Accuracy improves when more distractors are present. *Journal of Vision, 3,* 877–892.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 411–422). Cambridge, England: Cambridge University Press.

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception, 30,* 1285–1294.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–108.

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1194–1209.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In IEEE *Proceedings of the International Conference on Image Processing* (Vol. 1, pp. 253–256). Los Alamitos, CA: IEEE Computer Society.

O'Toole, A., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America, Series A, 10,* 405–411.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. Wenger & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 291–303.

Petrov, Y., & Zhaoping, L. (2003). Local correlations, information redundancy, and sufficient pixel depth in natural images. *Journal of the Optical Society of America A, 20,* 56–66.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–293.

Postma, E. O., Van den Herik, H. J., & Hudson, P. T. W. (1997). SCAN: A scalable neural model of covert attention. *Neural Networks, 10,* 993–1015.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002). Visual search: Structure from noise. In *Proceedings of the Eye Tracking Research & Applications Symposium* (pp. 119–123). New York: Association for Computing Machinery.

Rao, R. P. N., & Ballard, D. H. (1995). An active vision architecture based on iconic representations. *Artificial Intelligence, 78,* 461–505.

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

Steyvers, M. (2000). *Modeling semantic and orthographic similarity effects on memory for individual words.* Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.

Steyvers, M., & Busey, T. (2000). Predicting similarity ratings to faces using physical descriptions. In M. Wenger & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 115–146). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. Healy (Ed.), *Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* Washington, DC: American Psychological Association.

Tenenbaum, J. B., Silva, V. de, & Langford, J. C. (2000, December 22). A global geometric framework for nonlinear dimensionality reduction. *Science, 290,* 2319–2323.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.

Valentine, T. (1991). Representation and process in face recognition. In R. Watt (Ed.), *Vision and visual dysfunction,* V*ol. 14: Pattern recognition by man and machine*. London, England: Macmillan.

Viken, R. J., Treat, T. A., Nosofsky, R. M., McFall, R. M., & Palmeri, T. J. (2002). Modeling individual differences in perceptual and attentional processes related to bulimic symptoms. *Journal of Abnormal Psychology, 111,* 598–609.

Wainwright, M. (1999). Visual adaptation as optimal information transmission. *Vision Research, 39,* 3960–3974.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1022–1041.