

2D or not 2D? that is the question: What can we learn from computational models operating on 2D representations of faces?

Dominique Valentin

Université de Bourgogne à Dijon

Hervé Abdi

The University of Texas at Dallas

Betty Edelman

The University of Texas at Dallas

Mette Posamentier

The University of Texas at Dallas

Introduction

Recent work in automatic face recognition indicates that the major problem in modeling face processing is to find a meaningful representation for the faces. The difficulty arises from the often noted fact that human faces constitute a set of highly similar objects. Hence, a first constraint imposed upon a facial representation is that it must capture the subtle variations in features and configurations of features that make one face different from all other faces. This constraint makes an object-centered representation, such as the structural representation proposed by Marr and Nishihara (1978), or the geon-based representation proposed by Biederman (1987), an improbable candidate. While this type of representation seems to be appropriate

Thanks are due to John Vokey, Tom Busey, and the editors of this volume for helpful comments on earlier drafts of this paper.

Correspondence should be sent to: Dominique Valentin, ENSBANA, 1, Esplanade Erasme, Campus Universitaire, 21000 Dijon, France, E-mail: valentin@u-bourgogne.fr; <http://www.u-bourgogne.fr/d.valentin>, or Hervé Abdi, School of Human Development, The University of Texas at Dallas, MS:Gr4.1, Richardson, TX 75083-0688, E-mail: herve@utdallas.edu; <http://www.utdallas.edu/~herve>.

Ref: Valentin, D., Abdi, H., Edelman, B., Posamentier, M. (2001). 2D or not 2D? that is the question: What can we learn from computational models operating on 2D representations of faces? in M. Wenger, J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition*. Mahwah (NJ): Erlbaum. pp. ***-***.

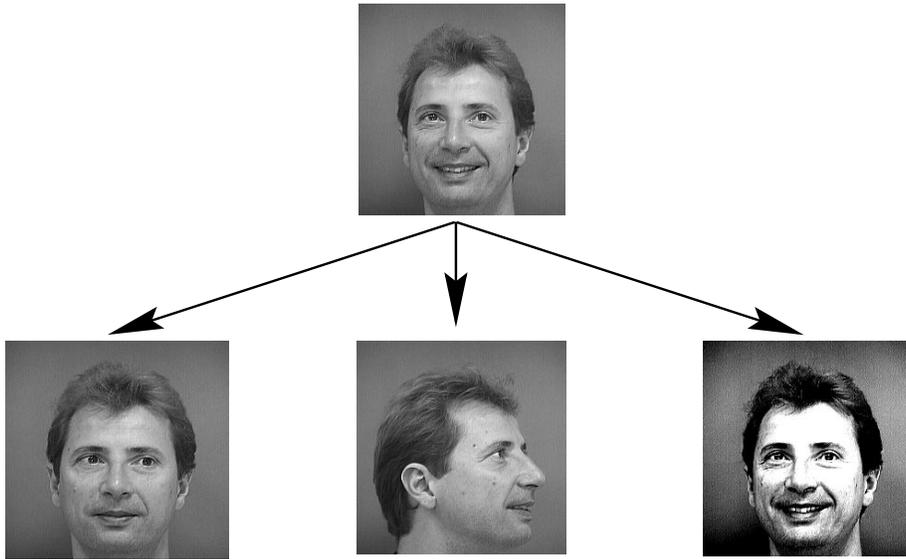


Figure 1. Illustration of the effect of expression, orientation, or lighting on the perceptual appearance of a face.

for assigning objects to basic level categories (Rosch, 1978), its extension to faces is problematic: broadly similar 3D structures would be obtained for all faces and no discrimination would be possible. The problem of quantifying similarity is explored in greater depth in the chapters of O'Toole, Wenger and Townsend; and Steyners and Busey, this volume.

An additional difficulty comes from the fact that the perceptual appearance of a face changes dramatically with changes in expression, orientation, or lighting (*cf.* Figure 1). Hence, a second constraint imposed upon a face representation is that it should be flexible enough to accommodate these transformations. Early models of face processing attempted to solve this problem by using a geometrical coding of faces. Key features were localized in the faces, and various measurements taken between these key features. This type of coding has the advantages of (a) capturing information useful for discriminating among faces, and (b) being relatively insensitive to transformations. Its main drawback is that it discards texture information that might be useful for tasks such as sex or race categorization or identifying the expression or the orientation of a face.

To preserve the texture information useful for performing categorization tasks, most of the recent computational models of face processing operate directly on an image-based coding of the faces (*i.e.*, a face is represented as a 2D array of pixel intensities). The main problem, however, with this type of representation is that it is not inherently 3D invariant. As a consequence, most of the current models using this type of coding are not able to handle changes in facial expression, lighting, or orientation (Turk & Pentland, 1991). However, it should be noted that these models

generally operate on a single frontal view of the face. A first question, therefore is: Can this limitation of 2D pixel based representations be overcome by using a set of 2D views sampling the rotation of the head from frontal to profile views? A second question is: How much information can be transferred from a single view of a face?

The main purpose of the work presented here is to address these two questions. After a brief presentation of autoassociative memories, we present recent data showing that a linear autoassociative memory trained to reconstruct multiple views of faces is able to generalize to new views of the faces. We then examine further the ability of the memory to transfer information from single views of faces, and compare the performance of the memory with that of human subjects on a similar task. Finally, we discuss our results in terms of a dual strategy approach to processing face images, and suggest a possible way of modeling such a dual process.

Autoassociative memory

Overview

In this section we provide an intuitive overview of the *autoassociative memory model*, also called the *autoassociator* or sometimes, in the face literature, the principal component (PCA) model. A more formal presentation can be found in several sources, our own favorite ones being (surprisingly?) Abdi (1994a and b), or Abdi, Valentin, and Edelman (1999, and in press). A mathematical *précis* can also be found in Appendix A.

An autoassociative memory is a neural network model in which the association between an input pattern and itself is learned. An important property of the autoassociative memory is to act as a pattern completion device because it is capable of reconstructing learned patterns when noisy or incomplete versions of these patterns are used as “memory keys.” From a statistical point of view, storing patterns in an autoassociative memory is equivalent to performing a principal component analysis of the set of faces (Abdi, 1988). In this framework, the principal components, often referred to as *eigenfaces* (Turk & Pentland, 1991), are interpreted as macro-features describing the faces.

Since Kohonen (1977) first demonstrated that an autoassociative memory can be used as a content addressable memory for face images, autoassociative memories have been successfully applied to the problems of face recognition (Millward & O’Toole, 1986), and categorization along visually based dimensions such as sex or race (Abdi, Valentin, Edelman, & O’Toole, 1995; Edelman, Valentin, & Abdi, 1998; O’Toole, Abdi, Deffenbacher, & Valentin, 1993; Valentin, Abdi, Edelman, & O’Toole, 1997). Although not intended as a general solution to the problem of face processing, autoassociative memories do provide a way of simulating some well-known psychological phenomena such as the other-race effect (O’Toole, Deffenbacher, Abdi, & Bartlett, 1991), the effect of typicality (O’Toole, Abdi, Deffenbacher, & Valentin, 1995), and the 3/4 view advantage (*i.e.*, 3/4 views are better recognized than either full face or

profile views, see Valentin, Abdi, & Edelman, 1997). The main problem of this type of approach is that, contrary to human observers, an autoassociative memory trained on face images is quite sensitive to changes in size, background, and to a lesser degree, lighting condition. The size and background problem, however, can easily be solved by automatically detecting the outline of the face in the picture and re-scaling it prior to recognition testing (Turk & Pentland, 1991).

What is an autoassociator?

To store a face in an autoassociative memory, the face image is first captured as grey-scale (or digitized) and transformed into a vector (called a *face vector*) by concatenating the columns of the corresponding image. A face vector element gives the value of the gray level of the corresponding pixel of the face image (this step is described in Figure 2a). The magnitude of the vector is, for convenience, normalized (*i.e.*, its length, or magnitude becomes 1). This has the effect of controlling for the overall illumination of the face image.

Next, each element of the face vector is used as input to a cell (or linear unit) of the autoassociative memory. The number of cells of the memory is equal to the number of elements in the vector, and each element of the vector is associated to one and only one cell. The cells are linked to each other by weighted connections. To store a pattern, the level of activation of each cell is set (*i.e.*, “clamped”) to the value of the corresponding image pixel. The cell then propagates its activation to all the other cells through the weighted inter-cell connections. This is illustrated by Figure 2b.

The “response” of the memory is obtained by letting each cell compute its new level of activation as the weighted sum of the activation of the other cells: The weights being given by the connections between cells. The response can be visualized by creating an image in which the gray level of each pixel is proportional to the activation of the corresponding cell. This is illustrated by Figure 2c, which shows the answer of an autoassociative memory prompted with the face of “Toto.” Supposing that only smiling faces were learned, the memory displays its pattern completion property when responding to a new view of “grumpy Toto” by giving back a smiling face (presumably a “smiling Toto”).

The quality of recall is evaluated by comparing input and output. The more similar input and output are, the more likely the input was previously learned by the memory. Among the several ways of evaluating similarity between face images, the most direct one is visual inspection. A numerical index can be computed as the squared coefficient of correlation between face vectors. This index varies between 0 (complete independence) and 1 (identical images). A variation of the coefficient of correlation is the cosine between face vectors (for all practical purposes, these two indices can be considered equivalent).

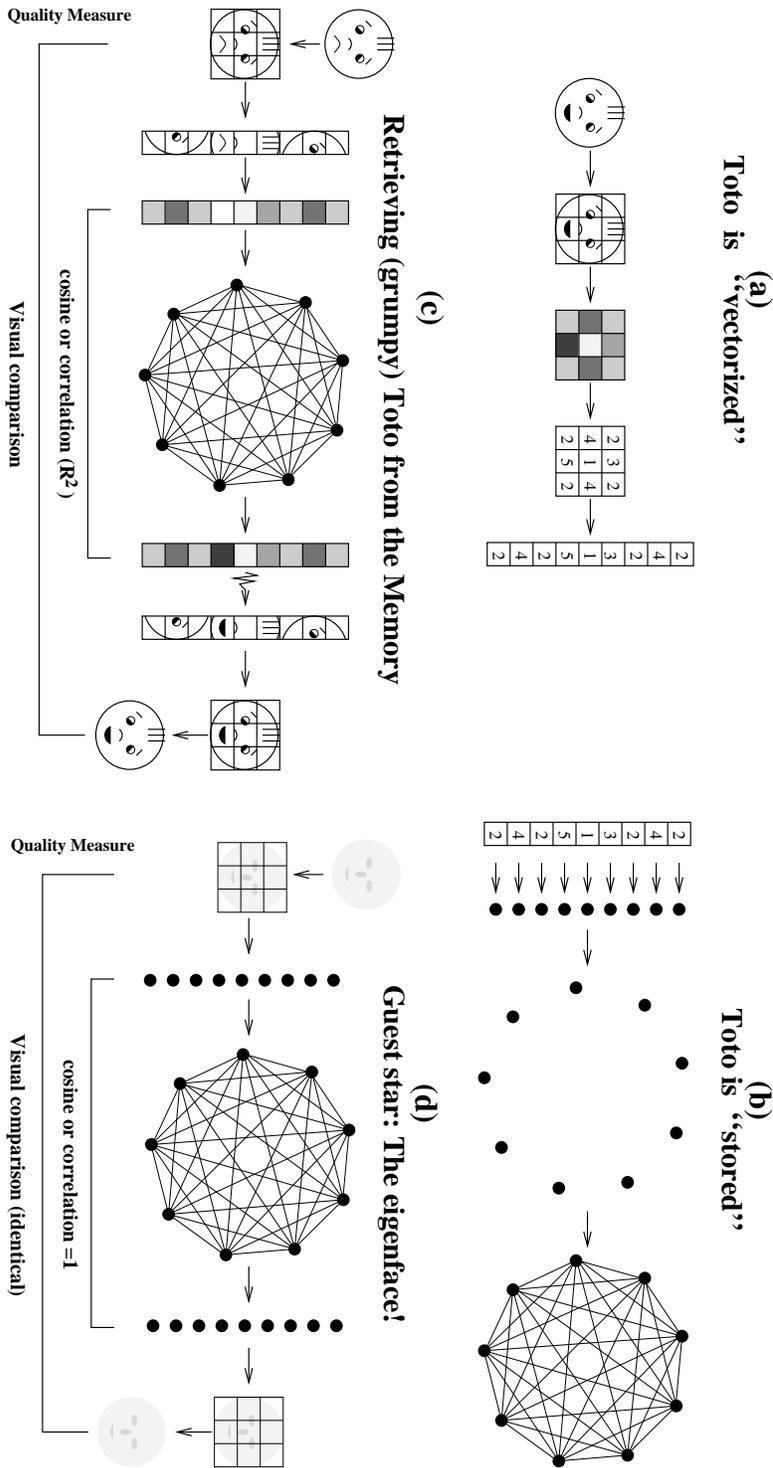


Figure 2. The major steps of an Autoassociative memory. (a) Step 1. A face image is transformed into a face vector. (b) Step 2. To each pixel of the face vector corresponds a cell. Cells communicate through weighted connections. (c) Step 3. The answer of the memory to a face is obtained by clamping each unit to the gray value of the corresponding pixel. Then each unit computes its activation. An image is obtained by displaying the activation of the units as gray values. (d) Eigenfaces: The response of the matrix to an eigenface (*i.e.*, an eigenvector of a face matrix) looks the same as the input.

Guest star: The eigenface!

The similarity between input and output patterns is used as an index of familiarity. Cast into a psychological framework, a correlation of 1 between input and output is equivalent to being certain that the input is known. If, in fact, the input was not learned, a correlation of 1 corresponds to making a false alarm with absolute confidence. For a given set of connection weights, it is possible to find these patterns. They are called the *eigenvectors* (from the German *eigen* meaning characteristic or specific) of the weight matrix (*cf.* Figure 2d). In the face literature, they are often labeled *eigenfaces*. In brief, an eigenvector has the property that the response of the matrix is proportional to the input. The coefficient of proportionality is called the *eigenvalue* associated to the eigenvector. Eigenvectors and eigenvalues constitute the major tools for analyzing the linear autoassociator. The formal notion of eigenvector corresponds to the psychological concepts of prototype and macrofeatures. An eigenvector can be seen as a prototype, because, like a prototype, it corresponds to a best form abstracted from the data, and therefore creates a maximal false alarm. Eigenvectors can also be seen as *macrofeatures* because they can be used to build back the learned faces (*i.e.*, any learned face can be reconstructed as a weighted sum of eigenvectors as illustrated in Figure 3). Figure 7 shows some eigenfaces.

$$\begin{array}{rcccl}
 \text{Toto} & = & \text{Some of} & + & \text{Some of} \\
 & & \text{eigenface 1} & & \text{eigenface 2} \\
 \img alt="A black and white line drawing of a face with a neutral expression, representing the 'Toto' face." data-bbox="218 538 292 594"/> & = & .8 \times \img alt="A grayscale image of a face with a neutral expression, representing 'Some of eigenface 1'." data-bbox="428 538 502 594"/> & + & 1.5 \times \img alt="A cyan-colored image of a face with a neutral expression, representing 'Some of eigenface 2'." data-bbox="708 538 782 594"/>
\end{array}$$

Figure 3. Toto is reconstructed as a weighted sum of eigenfaces.

Learning

The autoassociative memory has two main ways of learning: Hebbian and Widrow-Hoff. Both are iterative procedures, which means that they process one “stimulus-response” at a time. Each processing step entails small modifications of the set of connection weights.

Hebbian learning sets the connection weights by increasing the value of the connection between cells every time the corresponding pixels are in the same state, and decreasing the value of the connection every time the corresponding pixels are in different states. As a consequence, after learning, the connections reflect the covariations between pixels. Widrow-Hoff is a more sophisticated rule, which essentially learns by taking into account the difference between the actual answer of the memory and the target answer (*i.e.*, the to-be-learned stimulus). The connection weights

are modified such that the magnitude of this difference will be smaller for a second (immediate) presentation of the same stimulus. After a learning period long enough, and if the learning parameters were properly chosen, Widrow-Hoff learning will find a set of weights that minimizes the sum of the squared errors for the set of faces learned.

Transfer from multiple views

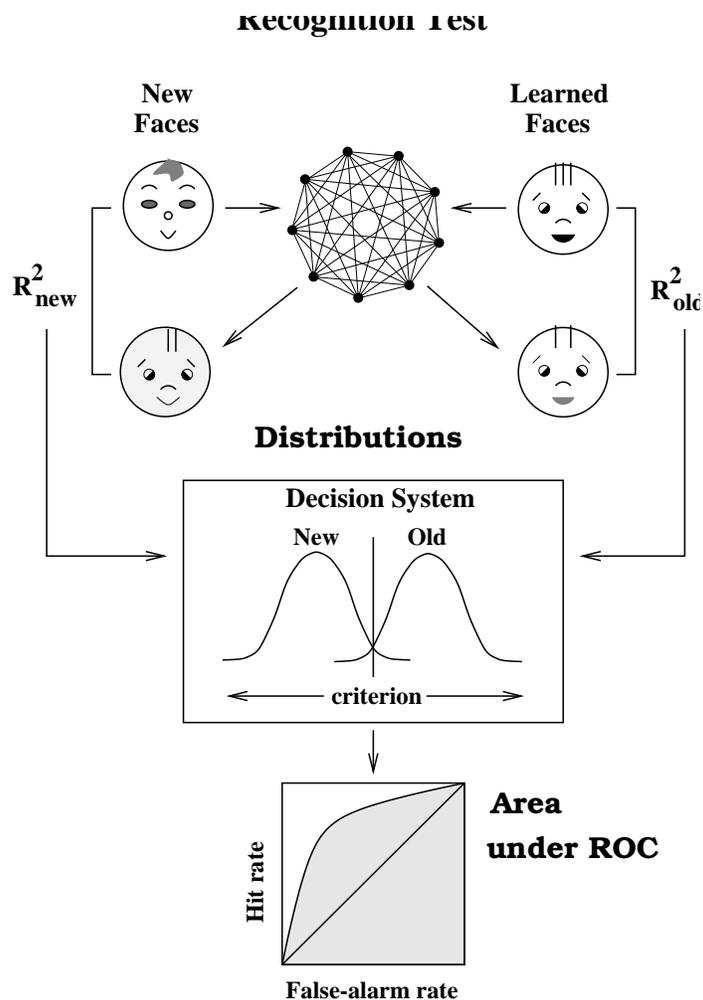


Figure 4. Illustration of the Valentin and Abdi (1996) testing paradigm.

Valentin and Abdi (1996) examined whether a set of 2D representations provides enough information for an autoassociative memory to recognize faces from new orientations. They stored 15 target faces in an autoassociative memory using complete Widrow-Hoff learning. The faces were represented by either a single, four, or nine views sampling the rotation of the head from full-face to profile. After learning completion (*i.e.*, when all the images were perfectly reconstructed), new views of

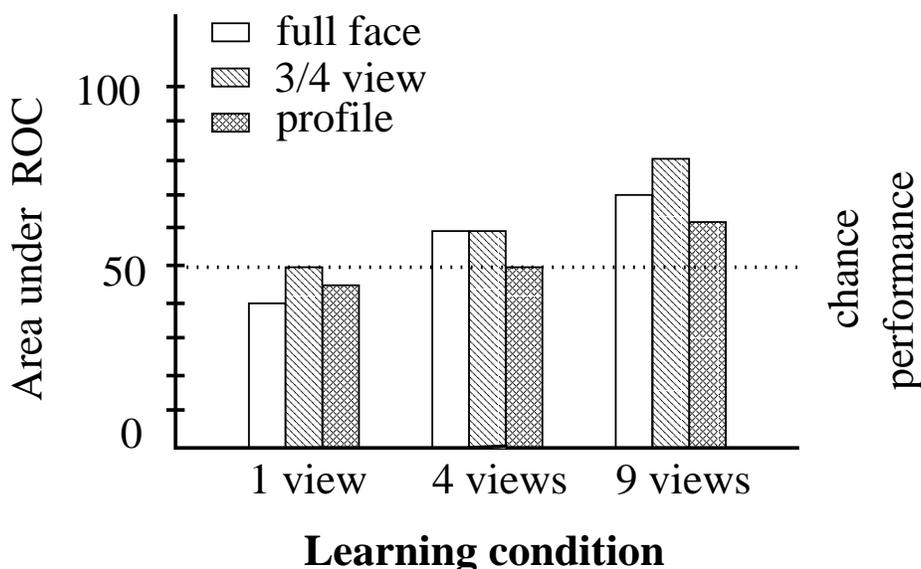


Figure 5. Transfer performance of the autoassociative memory model: Area under ROC as a function of learning conditions and view angles. The white bars represent the performance when a full-face was presented at test, the striped bars the performance when a 3/4 view was presented at test, and the cross-hatched bars the performance when a profile was presented at test.

the target faces and an equal number of distractor faces were presented as input to the memory. For each view, a cosine was computed between the original and reconstructed images. The cosine provides an indication of the familiarity of the model with the faces. The higher the cosine is, the more probable it is that the face has been learned. The recognition task was implemented by setting a criterion cosine and by categorizing each face with a cosine greater than the criterion as “old” or learned and each face with a cosine smaller than the criterion as “new” (*cf.* Figure 4). Different criteria were used so as to generate a receiver operating characteristic (ROC) for each learning condition. The area under the curve (*i.e.*, gray area in Figure 4) provides an unbiased estimate of the proportion of correct classification, with a chance level at 50% (Green and Swets, 1966).

The results are summarized in Figure 5. In the 1-view condition, the memory is not able to generalize to a new view of a learned face, no matter which view is presented at test (area under ROC $\approx .5$). In the 4-view condition, with either a frontal or a 3/4 test view, the performance of the memory is somewhat better than in the 1-view condition. In the 9-view condition, the memory is clearly able to discriminate between old and new faces. In other words, when an autoassociative memory is made of single views of faces, its ability to recognize faces from new view angles is somewhat

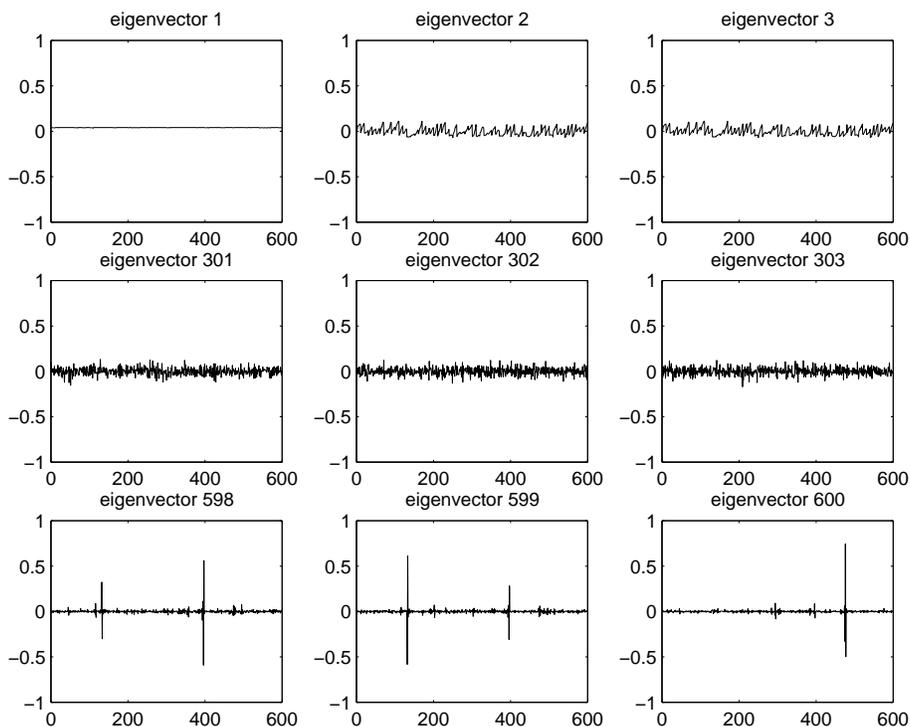


Figure 6. Normalized projections of 60 faces, each represented by 10 views (for a total of 600 images) sampling the rotation in depth from full-face to profile, onto different eigenvectors. The horizontal axes represent the face images. The first 10 images represent the 10 views of the first face, the next 10 images the 10 views of the second face, *etc.* The vertical axes represent the projections, or coordinates, of the faces onto the eigenvectors. The projection of a face onto an eigenvector indicates the importance of the face for the eigenvector. The higher the value is, the more the face contributes to the eigenvector.

similar to that observed for human subjects presented with unfamiliar faces (*i.e.*, it is sensitive to depth rotation). In contrast, when the memory is trained with multiple views sampling the rotation of the head from full-face to profile, its performance parallels that of human subjects presented with familiar faces (*i.e.*, it becomes less sensitive to depth rotation).

Using an autoassociative memory as a content addressable memory for faces is equivalent to performing a principal component analysis on a set of face images (Abdi, 1988). Thus, standard statistical techniques can be used to interpret and represent the information contained in the set of faces (O’Toole, Abdi, Deffenbacher, & Valentin, 1993). For example, Figure 6 shows the projections of 600 face images (60 faces \times 10 views) onto (a) the three eigenvectors with the largest eigenvalues, (b) three eigenvectors with intermediary eigenvalues, and (c) the three eigenvectors with the smallest eigenvalues of the set of faces. This figure shows that different ranges of eigenvectors convey different types of information. Specifically:

- Eigenvectors with large eigenvalues contain information relative to the orientation and general shape of the faces. These eigenvectors are useful in detecting the particular orientation of faces.
- Eigenvectors with intermediate eigenvalues contain information specific to small sets of faces across orientations. These eigenvectors are useful in interpolating between views of particular faces.
- Eigenvectors with small eigenvalues contain information relative to the identity of the faces. These eigenvectors are useful for discriminating between faces.

An examination of individual eigenvectors confirms this dissociation of orientation and identity-specific information. As an illustration, Figure 7 displays the first three and the last three eigenvectors of the autoassociative memory. Clearly, the last three eigenvectors are specific to particular faces in a particular orientation. In contrast, the first three eigenvectors capture information that is common to many faces. The first eigenvector represents a kind of average across faces and orientations. The second and third eigenvectors oppose profile to frontal views for all the faces.

This dissociation is reminiscent of some physiological data. Perrett, Rolls, and Caan (1982) found a population of cells in the fundus of the superior temporal sulcus of three rhesus monkeys that were selectively responsive to human and monkey faces. These “face selective cells” responded to many different faces but were able to cope only with limited depth rotations. Rotating the faces from full-face to profile reduced or eliminated the response of 60% of the cells. Even rotations as small as 10 or 20 degrees produced a substantial reduction of the responses. In addition to the cells tuned to specific views, Perrett *et al.* (1986) reported finding some cells, or groups of cells, responding to specific faces across different viewing orientations. Hasselmo, Rolls, Baylis, and Nalwa (1989) found 37 face-selective neurons in the superior temporal cortex of three macaque monkeys. Consistent with the findings of Perrett *et al.* (1982, 1986), they reported that, out of these 37 cells, 18 showed selectivity for specific faces across different view angles. Out of the 19 remaining neurons, 16 showed selectivity for specific views across the faces. In addition, they reported that, among the 18 identity-specific neurons, 15 showed some response modulations as a function of the viewing angle.

The coexistence of *view-independent* and *view-dependent* neurons in the superior temporal sulcus has been interpreted by Hasselmo *et al.* (1989) as an indication that object-centered representations are built from different views of the faces in this area. This coexistence can also be interpreted as evidence for the existence of two kinds of facial information simultaneously stored in memory. The first one, view-independent, would be useful to identify a particular face and the second one, view-specific, would be useful to remember episodic information about the face. According to Bruce (1982), both view-independent (*i.e.*, structural) and view-dependent (*i.e.*, pictorial) information would be stored for familiar faces, but only view-dependent information would be stored for unfamiliar faces.

In conclusion, Valentin and Abdi’s results suggest that faces might be repre-

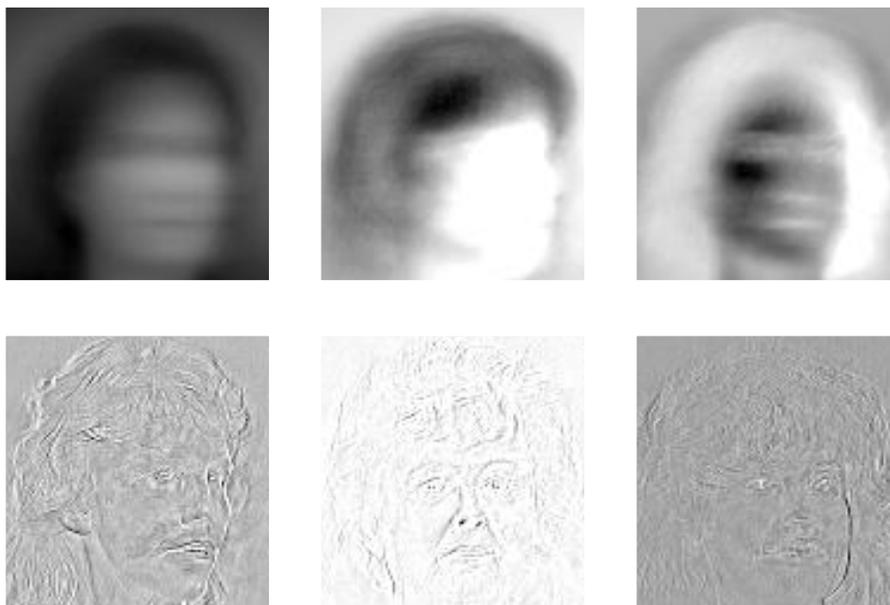


Figure 7. The first three and the last three eigenvectors of a face autoassociative memory trained to reconstruct 40 faces each represented by 10 views sampling the rotation of the head from full-face to profile.

sented in memory using a set of 2D view-dependent representations. The only constraint is that the learned views need to be spaced closely enough for the model to use interpolation between two views in order to transfer to new views. The remaining question is: How close should the views be? To answer this question we analyzed further the ability of both human observers and an autoassociative memory to transfer information from single views of faces.

Transfer from single view

Experiment 1

Over the past 20 years several studies have shown that human observers' recognition performance is significantly impaired when faces are rotated in depth between learning and test. However, because these studies used only a small number of rotations (0, 45, and 90 degrees), they were not able to determine how much faces can be rotated in depth between learning and test before impairing recognition performance. The purpose of the present experiment was to extend the scope of the previous studies by using eight rotation conditions (0, 10, 20, 30, 40, 50, 60, and 90) between learning and test. This approach provides the data for a fine-grained analysis of the ability of human observers to transfer information from single views of faces.

Methods

Observers.

Sixty-four undergraduate students from the School of Human Development of the University of Texas at Dallas (UTD) were recruited in exchange for a core psychology course research credit. The fact that they were not familiar with the faces was verified at the end of the experiment, and only the data obtained from observers unfamiliar with the faces were analyzed.

Stimuli. Forty Caucasian female graduate students, staff and faculty members of the School of Human Development participated in the creation of a database. Twenty images *per* person were captured by a video camera and digitized with a resolution of 256 gray levels by a RasterOps 24STV board connected to an Apple Macintosh Quadra 610. The 20 views included one series of ten views sampling the rotation of the head from full-face to right profile with about 10-degree steps (*i.e.*, 0, 10, 20, 30, 40, 50, 60, 70, 80, and 90 degrees from the camera) and two series of five views, each sampling the rotation of the head from full-face to right profile with about 20-degree steps (*i.e.*, 0, 20, 50, 70, and 90 degrees from the camera).

The collection of the images was done as follows. Each person was seated in front of the video camera and asked to rotate her head in progressive steps from full-face to right profile while keeping the same neutral expression. To ensure that all the faces were taken at roughly the same angle of rotation, red vertical lines were drawn on the wall beside the video camera to indicate the different angles at which the images were to be captured. After facing the camera directly (frontal view), each person was instructed to point her nose at each of these lines, in turn. The lighting conditions were the same for every person. None of the captured face images showed any major distinguishing characteristics, such as jewelry, glasses, or clothing. All the images were roughly aligned along the axis of the eyes so that the eyes of all faces were at the same height. The final images were 230 pixels wide and 240 pixels high.

Thirty faces were selected from the database to be used in turn as targets or distractors. The ten remaining faces were used as fillers during the learning session.

Experimental design. The observers were tested on a standard yes-no recognition task. A one-factor between-subject factorial design was used with *angle of rotation* between learning and test (0, 10, 20, 30, 40, 50, 60, 90 degrees) as the independent variable and *recognition accuracy*, *decision bias*, and *reaction time* as dependent variables.

A counterbalancing procedure was used to ensure that every face appeared equally often as target and distractor. For both learning and testing, the order of presentation of the faces was randomized and a different order used for each observer. In the 0-degree condition, different pictures representing the same orientation (*e.g.*, two different images of a profile view) of the target faces were used for learning and testing.

Apparatus. Experimentations were performed with programs written in the C language on a Sun SparcStation 5 running X11 under UNIX.

Procedure. The experiment consisted of two sessions, learning and testing, separated

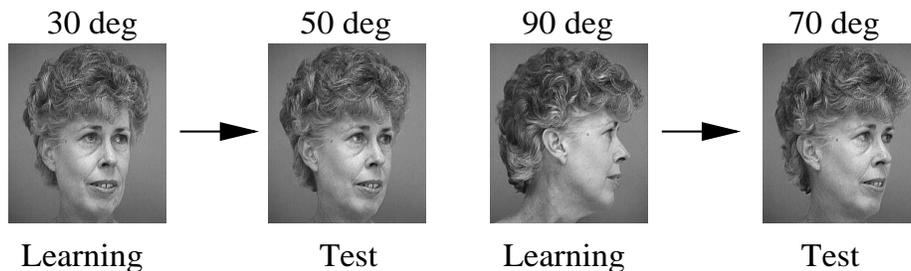


Figure 8. Examples of 20-degree rotation conditions.

by a 10-minute break. During the learning phase, observers were shown 25 faces (15 targets and 10 fillers), each presented on a computer screen for 4 seconds, with a 4-second inter-stimulus interval. An approximately equal number of faces appeared in each of 10 possible orientations¹ (0, 10, 20, 30, 40, 50, 60, 70, 80 and 90-degree rotation from a frontal view). Observers were asked to watch the faces and to try to memorize them. They were informed that a recognition test would follow, and that the faces at test would not necessarily be in the same orientation as originally presented. During the testing phase, observers were shown a series of 30 faces (the 15 targets mixed with 15 distractors). For 1/8 of the observers, the target faces were in the same orientation as during learning. For the remaining observers, the target faces were rotated in depth with either a 10, 20, 30, 40, 50, 60, or 90-degree rotation between learning and testing. Figure 8 displays some examples of rotations used in the experiment. Note from this figure that any given rotation condition can be obtained with different pairs of views. For example, a 20-degree rotation can be obtained by presenting a target face in a 30-degree orientation during learning and in a 50-degree orientation during testing, or by presenting it in a 90-degree orientation during learning and in a 70-degree orientation during testing.

For each face, observers were instructed to press the right mouse button if they thought the face was presented during the learning session and to press the left mouse button if they thought it was not presented during the learning session. The faces remained on the computer screen until the observers indicated their answer by pressing one of the mouse buttons. As reaction time was recorded, observers were asked to answer as accurately and as quickly as possible.

Results.

Results were analyzed using signal detection methodology. Each observer contributed an accuracy index ($d' = z_{\text{hit}} - z_{\text{false-alarm}}$) and a bias index [$C = -\frac{1}{2}(z_{\text{hit}} + z_{\text{false-alarm}})$] calculated on the basis of 15 scores². Hit rates of 100 per-

¹In the 0-degree condition, we used only the orientations for which two different images were available (*i.e.*, 0, 20, 50, 70, and 90-degree rotation from a frontal view).

²Note that the formula used to compute d' depends upon the way the z -transform is formalized. As a consequence, different authors (*e.g.*, Mc Nicol, 1972) will give apparently different formulas.

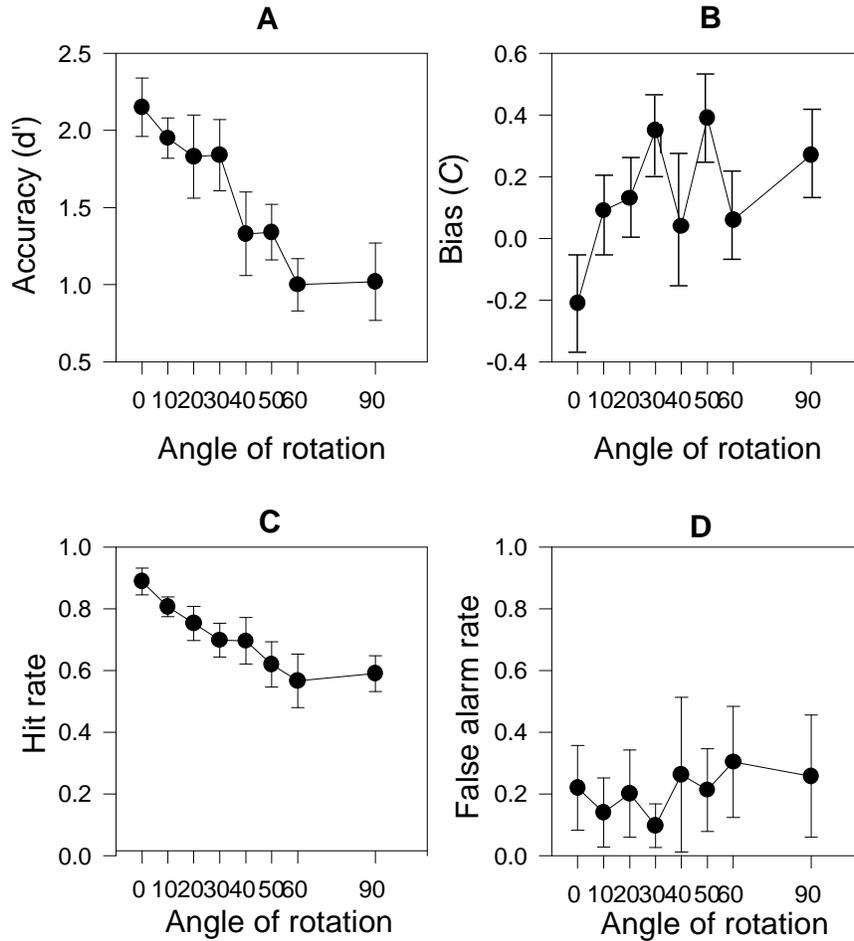


Figure 9. d 's (A), C s (B), hits (C), and false alarms (D) as a function of the rotation angle between learning and tests. Error bars show the standard error of the mean.

cent and false-alarm rates of 0 percent were converted to $1 - \frac{1}{2N} = .97$ and $\frac{1}{2N} = .03$, respectively, with N representing the number of scores (*cf.* Macmillan & Creelman, 1991), thus leading to a maximum value of d' equal to 3.76. Separate one-factor between-subject ANOVAs³ were carried out for d' and C . An additional analysis was then carried out to examine the pattern of response times recorded for correct responses to target faces (hits).

Recognition accuracy. The mean d' values are shown in Figure 9A. The ANOVA

But, indeed, *mutatis mutandis*, the result is always the same.

³Using analysis of variance with d' carries the potential problem of violation of the assumptions of normality and homogeneity of variance. When the design is balanced and when the number of degrees of freedom is large enough (both conditions being fulfilled here), however, this problem is of no practical consequences (see, *e.g.*, Abdi, 1987, p. 128 *ff.*).

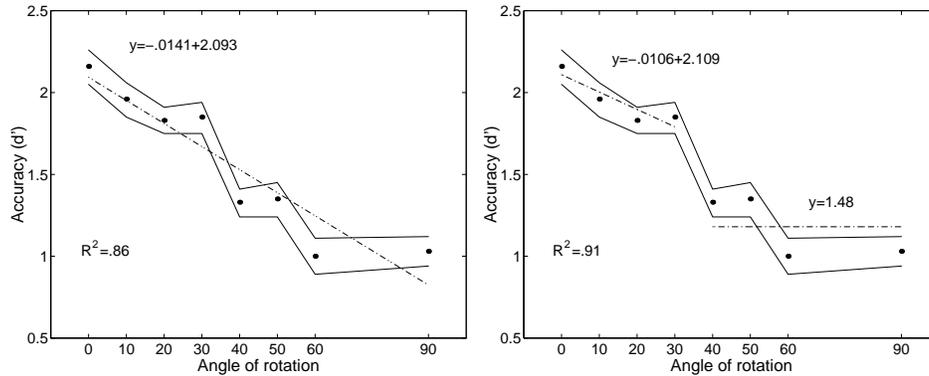


Figure 10. Linear regression of the average d' s as a function of angle of rotation. The black circles represent the average d' s, the solid lines the confidence interval, and the dashed lines the regression lines. The left panel represents a single-linear-function regression. The right panel represents a two-linear-function regression. The best fit is obtained with the two-function regression.

reveals a significant effect of *angle of rotation*, $F(7, 56) = 3.96$, $MS_e = .39$, $p < .01$. Performance accuracy decreases as an inverse function of the angle of rotation between learning and test from 2.15 (when no change occurs between learning and test) to 1.02 (after 90 degrees). Figures 9C and D show that this decrease is due to a clear diminution of hits and a slight augmentation of false alarms.

A trend analysis, shows that, only 30% of the *total variance* in accuracy performance can be explained by the existence of a linear relationship between angle of rotation and recognition accuracy. Moreover, a series of linear regression analyses (Figure 10) carried out on the average d' s shows that a two-function regression provides a better fit ($R^2 = .91$) than a one-function regression ($R^2 = .86$). Finally, pairwise comparisons using a Duncan test show that the rotation conditions fall into two groups (0, 10, 20, and 30, on one hand, and 40, 50, 60, and 90, on the other hand), which do not vary significantly within themselves. These results suggest that human observers are able to cope with up to 30 degrees depth rotation ($d' = 1.94$, on the average), but after this cut-off point, a significant decrement in performance is observed ($d' = 1.17$, on the average).

Decision bias. The mean C values are shown in Figure 9B. The ANOVA fails to reveal a global effect of angle of rotation on decision bias, $F(7, 56) = 1.56$, $MS_e = .19$, $p > .05$. However, pairwise comparisons using a Duncan test reveal a significant difference in the decision criterion used by the observers when no change occurred between learning and test and after either a 30- or a 50-degree rotation. Observers tend to be liberal (*i.e.*, they tend to say yes) in the 0 condition and conservative (*i.e.*, they tend to say no) in the 30, 50, and to a lesser extent 90 degree conditions. In the

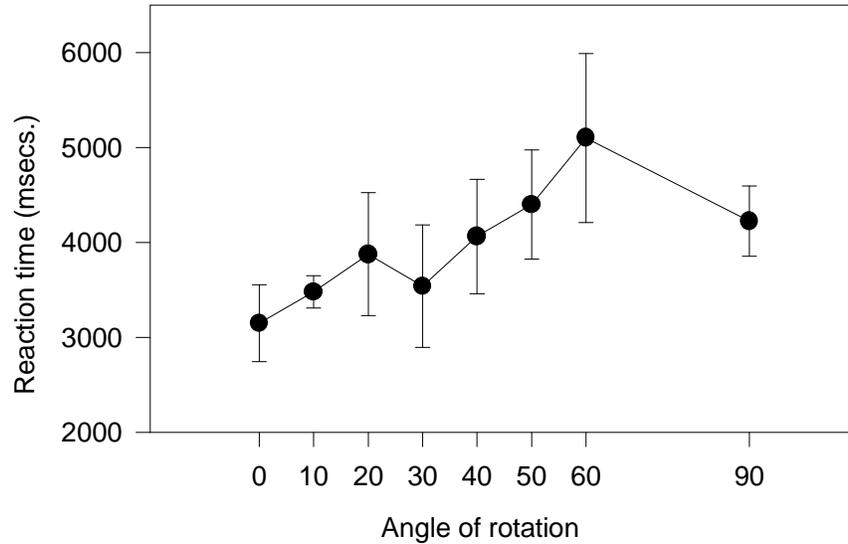


Figure 11. Reaction times for correct recognition averaged across observers as a function of the rotation angle between learning and testing. Error bars show the standard error of the mean.

other conditions, observers tend to be neutral (bias close to zero).

Reaction time. The median of each observer's reaction time distribution for hit was calculated across faces and used as a dependent variable for the following analyses. Figure 11 shows the average reaction times as a function of the angle of rotation between learning and test. The data from one outlier observer in Condition 60-degree (further than 2 standard deviations from the average) have been discarded for this analysis. This figure shows that, except for a small dip at 30 degrees, reaction time increases as a function of the rotation angle up to 60 degrees and decreases from 60 to 90 degrees. However, a trend analysis shows that only 10% of the *total variance* of reaction times can be explained by the existence of a linear relationship between angle of rotation and reaction time. Pairwise comparisons using a Duncan test indicate that the difference between 60 and 90 degrees is significant ($\alpha = .05$).

Discussion.

The results of this experiment replicate the finding that human observers' recognition performance for unfamiliar faces is affected by a change in rotation between learning and test. Precisely, it was found that the ability of observers to recognize a face previously seen from a single viewpoint is stable up to 30 degrees, and deteriorates when the change between learning and test involves a depth rotation greater than 30 degrees. It should be noted, however, that, although observers become sig-

nificantly less accurate after 30 degrees, their performance is still above chance (d' reliably different from 0).

The patterns of results for d' , C , and reaction time obtained in this experiment provide some insight into the strategies or kind of information that might be used by human observers to solve this type of recognition task. First, the small amount of variance explained by a linear relationship between angle of rotation and accuracy, on one hand, and angle of rotation and reaction time on the other hand (30% and 10% respectively), along with the non-monotonic aspect of the curves presented in Figures 9 and 11, makes an interpretation in terms of mental rotation unlikely (*cf.* Shepard & Cooper, 1982). In this context, a mental rotation theory would predict that the reaction time necessary to mentally rotate the faces should increase as a *monotonic function* of the rotation angle applied between learning and test. The fact that the observed reaction time was significantly smaller when the faces were rotated by 90 degrees than when they were rotated by 60 degrees, thus, cannot be accounted for by such a theory.

Second, the drop in recognition accuracy observed after 30 degrees associated with the changes of criterion—first from liberal (0 degrees) to conservative (30 degrees), then in a more chaotic way from 40 degrees to 90 degrees—suggests that different transfer strategies might be used by human observers to recognize faces presented in new orientations, depending on the amount of transformation from the original image. Although it is not clear from these data which type of information is transferred in such tasks, we can speculate that, up to 30 degrees, observers tend to use global configural information associated with a matching strategy. Whereas this strategy is very efficient when no major changes occur between learning and test, it becomes less efficient when the difference between the original image and the image presented at test increases. From our data, it seems that up to 30 degrees, the similarity between the original image and the rotated image is large enough to allow such a configural transfer. However, after 30 degrees, the similarity between original and rotated images seems to be too small to allow this type of transfer. If this is the case, the same decline in performance after 30 degrees should be observed for an autoassociative memory, because the ability of the memory to recognize a face from a new orientation is based essentially on the existence of a correlation between the original view of the face and the transformed one. This conjecture will be tested in the following simulations.

After a 30-degree rotation, when no configural transfer is possible, observers seem able to transfer another type of information that is somewhat invariant to depth rotation⁴. Although this information is not as useful as the global configural information to perform the task, it allows for a recognition performance above chance level. The nature of this invariant information, however, is still unclear and will be reexamined later in light of the results obtained from the simulations and

⁴Note that this type of information could also, in some cases, be used to recognize faces after a rotation smaller than 30 degrees.

Experiment 2.

Simulations

The purpose of this series of simulations was to test the ability of an autoassociative memory to generalize to new views of faces learned from a single view. The performance of the memory was then compared with the performance of the human observers of Experiment 1. If our interpretation of the human data is correct, the performance of the model should also break down after a 30-degree rotation between learning and test. The method used was similar to that of Valentin and Abdi (1996) with the exception that a d' measure was used to evaluate the performance of the memory instead of a ROC curve.

*Methods**Stimuli.*

The same 30 faces (15 targets and 15 distractors) and 10 fillers as in Experiment 1 were used as stimuli.

Experimental design. As in Experiment 1, a one-factor design was used with *angle of rotation* between learning and test (0, 10, 20, 30, 40, 50, 60, and 90 degrees) as the independent variable and *quality of reconstruction* and *recognition accuracy* as dependent variables.

Eight series of simulations were carried out, one for each rotation condition. In each series, eight separate simulations were carried out so that every face appeared as target and distractor and in each transformation condition. As in Experiment 1, different pictures (in the same orientation) of the target faces were used for learning and testing in the 0-degree condition.

Procedure. This simulation included two phases, a learning phase and a testing phase, in which the ability of the memory to “recognize” learned faces from new views was tested. During the learning phase, 15 targets and 10 fillers were stored in an autoassociative memory using complete Widrow-Hoff learning. An approximately equal number of faces appeared in each of the 10 possible view orientations⁵ (0, 10, 20, 30, 40, 50, 60, 70, 80, and 90 degrees from full-face). At the end of learning, all the faces in the learning set were perfectly reconstructed by the memory (*i.e.*, the cosines between original and reconstructed faces were equal to 1). After learning completion, the weights of the autoassociative memory were fixed. Thirty faces (15 new images of the target faces and 15 distractors) were used as input to the memory. The new images of the targets were rotated in depth with either 0, 10, 20, 30, 40, 50, 60, or 90 degrees of rotation from the view presented during learning. For each face, targets and distractors, the quality of the response of the memory was evaluated by computing the cosine between reconstructed and original images. To simulate human observers’ decision procedure according to signal detection theory, we computed d' assuming the

⁵Again, only the orientations for which two images were available were used in the 0-degree condition (*i.e.*, 0, 20, 50, 70, and 90-degrees from a frontal view).

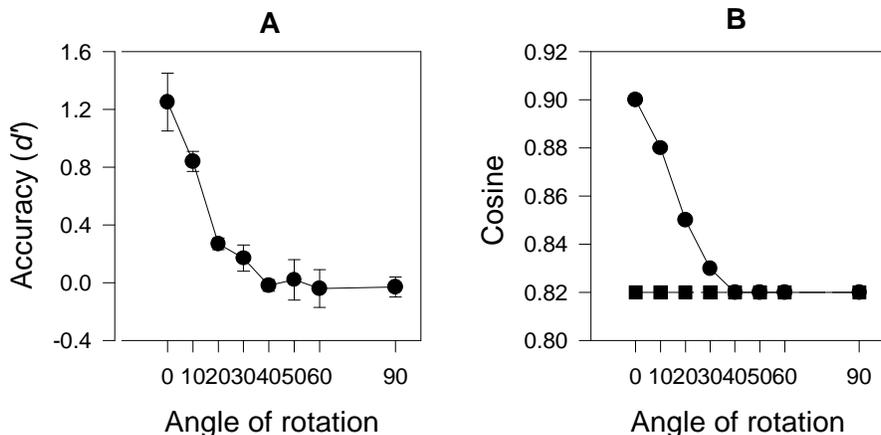


Figure 12. d' (left panel) and cosines (right panel) averaged across simulations as a function of the angle of rotation between learning and test. In the right panel, the squares represent the distractors and the circles the targets. Error bars show the standard error of the mean.

model behaves like the ideal observer. This was done by setting a criterion cosine (*i.e.*, the average of the target and distractor cosines), and by categorizing each face with a cosine greater than the criterion as “learned” and each face with a cosine smaller than the criterion as “new.”

Results.

Results were analyzed using standard signal detection methodology. For each simulation, a d' was calculated on the basis of 15 scores. A one-factor between-subject ANOVA was carried out with *angle of rotation* as the independent variable and *recognition accuracy (d')* as the dependent variable.

The mean d' values are shown as a function of the angle of rotation between learning and testing in Figure 12A. From this figure, it appears that the performance of the memory decreases as a function of the difference between learned and test views, $F(7, 50) = 16.01$, $MS_e = 1.54$, $p < .0001$. Pairwise comparisons using a Duncan test reveal a significant decrease in performance between 0 and 10 degrees and between 10 and 20 degrees. After 30 degrees, the memory is not able to differentiate between learned and new faces ($d' \approx 0$). As an illustration, Figure 12B shows the average cosines for target and distractor faces as a function of the angle of rotation between learning and test. This figure shows that, up to 30 degrees, target faces are better reconstructed, on the average, than distractor faces. After 30 degrees, target faces are no longer better reconstructed than distractor faces, and therefore, the memory is unable to discriminate between these two classes of faces.

Discussion.

Two main points can be noted from this series of simulations. First, the maximum amount of depth rotation an autoassociative memory trained on single views of faces can handle is about 30 degrees. This cut-off point is similar to that observed for human observers in Experiment 1. Because the performance of the memory is directly based on the pixel correlation between the input image and the face images stored in the memory, this result provides some support for the idea that a configural transfer based on global perceptual similarity between learned and new views is possible up to only 30 degrees. The fact that the memory performs at chance level ($d' \approx 0$) beyond this point indicates that there is not enough common information between the test image and any of the learned images for the memory to reconstruct the face. The second conclusion we can draw from this simulation is that, unlike human observers in Experiment 1, the autoassociative memory does not extract invariant information from single views of faces. After 30 degrees, the memory performs at chance level, whereas human observers were still performing above chance ($d' \approx 1$).

A possible explanation for the human observers' advantage can be found in the difference of learning history between human observers and the autoassociative memory. Whereas, in both cases, the specific faces used as stimuli are unfamiliar, human observers have a large amount of experience with this particular class of stimuli, but the autoassociative memory does not. In other words, the difference in performance between human observers and the autoassociative model might be attributable to a difference in the level of *general familiarity* with Caucasian faces. This hypothesis is tested in Experiment 2 using human observers with a lower level of general familiarity with Caucasian faces than observers in Experiment 1.

Experiment 2

The purpose of this second experiment is to examine how observers' general familiarity with a class of faces affects the ability to extract invariant information from single views of faces. The rationale for this experiment is as follows: If the level of recognition performance observed in Experiment 1 in the 60 and 90-degree conditions is due to the fact that observers were extremely familiar with the class of faces (*i.e.*, they have encountered many Caucasian faces before the experiment), then a lower level of performance should be observed for observers unfamiliar with this class of faces. To test this hypothesis, we replicated Experiment 1 using Asian observers, recently established in the United States, who had only limited exposure to Caucasian faces during their childhood. If the general familiarity hypothesis holds true, Asian observers' performance should be more disrupted by a rotation of the faces between learning and test than that of Caucasian observers and closer to that of the autoassociative model. However, if the ability to extract invariant information is not linked to the general familiarity with the class of faces, Asian observers should not be more affected by depth rotation of these faces than Caucasian observers.

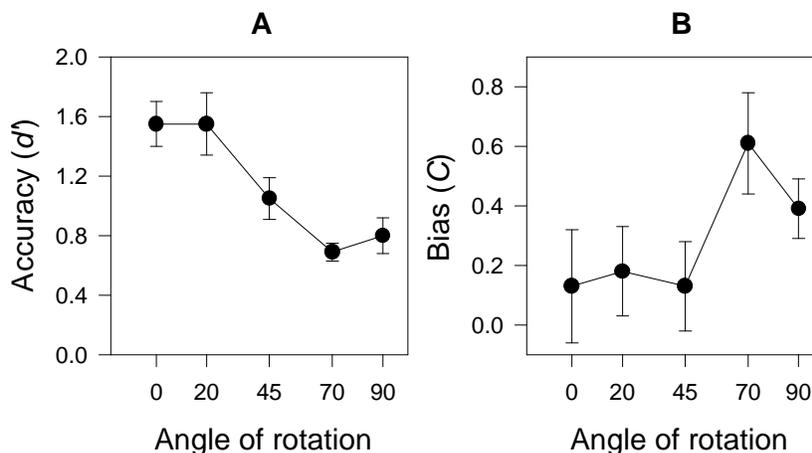


Figure 13. Average d' (left panel) and C (right panel) as a function of the rotation angle between learning and test. Error bars show the standard error of the mean.

Methods

Observers.

Thirty-four Asian observers (Korean, Chinese, and Vietnamese), living in the United States for less than five years, volunteered to take part in the experiment. None of the observers were familiar with the faces.

Stimuli. The same 30 faces and 10 fillers as in Experiment 1 were used as stimuli.

Experimental design. The design was the same as in Experiment 1 with the exception that because of the small number of observers available only five angles of rotation were used (0, 20, 50, 70, and 90 degrees) as the between-subject factor. In addition, the paucity of observers yielded a slightly unbalanced design with respectively 7, 7, 7, 6, and 7 observers *per* condition.

Procedure. The procedure was the same as in Experiment 1.

Results.

As for the previous experiment, results were analyzed using signal detection methodology. Each observer contributed a d' and a bias index C , calculated on the basis of 15 scores. Separate one-factor between-subject ANOVAs were carried out for d' and C .

Recognition accuracy. The mean d' values are shown in Figure 13A. The ANOVA reveals a significant effect of angle of rotation, $F(4, 29) = 6.11$, $MS_e = .16$, $p < .001$. As for Caucasian observers, performance decreases as an inverse function of the angle of rotation between learning and test—from 1.55 in the 0 and 20-degree conditions

to .80 in the 90-degree condition. Pairwise comparisons using a Duncan test show that accuracy performance fell into two groups corresponding to the 0 and 20-degree conditions on one hand, and the 50, 70, and 90-degree conditions on the other hand. The average d' in the first group is 1.55 and the average d' in the second group is .85, thus indicating a global decrement of .70. As a comparison, the global decrement in Experiment 1 (Caucasian observers) was .77.

Decision bias. The mean C values are shown in Figure 13B. The ANOVA shows a significant effect of angle of rotation, $F(4, 29) = 2.59$, $MS_e = .12$, $p < .05$. Observers tend to use a stricter criterion with a larger change in rotation. Pairwise comparisons using a Duncan test reveal a significant change in bias after 50 degrees: Before this cut-off point (*i.e.*, 0, 20 and 50-degree condition) the average value for C is .12, indicating a close to neutral decision bias. After this cut-off point (70 and 90-degree conditions), the average value for C is .50, indicating a more conservative decision bias.

Comparison between Caucasian and Asian observers.

An additional analysis was performed to compare the recognition accuracy of Caucasian and Asian observers. The data corresponding to the 0, 20, 50, 70, and 90-degree conditions of Experiment 1 were extracted and analyzed in conjunction with the data of Experiment 2. An unbalanced two-factor between-subject ANOVA with *race of observers* and *angle of rotation* as independent variables and d' as the dependent variable was carried out (using SAS PROC GLM Type III sums of squares). The results show:

- A main effect of *angle of rotation*, $F(4, 64) = 9.51$, $MS_e = 1.62$, $p < .0001$. On the whole, performance decreases when faces are rotated in depth between learning and testing.
- A main effect of *race of observers*, $F(1, 64) = 6.51$, $MS_e = 1.62$, $p < .01$. On the whole, Caucasian observers are more accurate than Asian observers ($d' = 1.46$ vs 1.14, respectively).
- No significant interaction between *angle of rotation* and *race*: Asian observers are no more affected by a rotation in depth between learning and test than Caucasian observers. Thus indicating that, in this experiment, rotation in depth is not more disruptive for observers with a low general familiarity with the class of faces than for observers with a high general familiarity.

Discussion.

The results of this experiment show that, indeed, human recognition performance is affected by the general familiarity with the class of faces. As expected, recognition accuracy decreases with the level of general familiarity with the faces. Asian observers are less accurate in a recognition task involving Caucasian faces than Caucasian observers are. But, contrary to an earlier finding by Ellis and Deregowsky (1981), this experiment fails to reveal an interaction between the angle of rotation and the race of observers. Asian observers recognizing Caucasian faces are not more

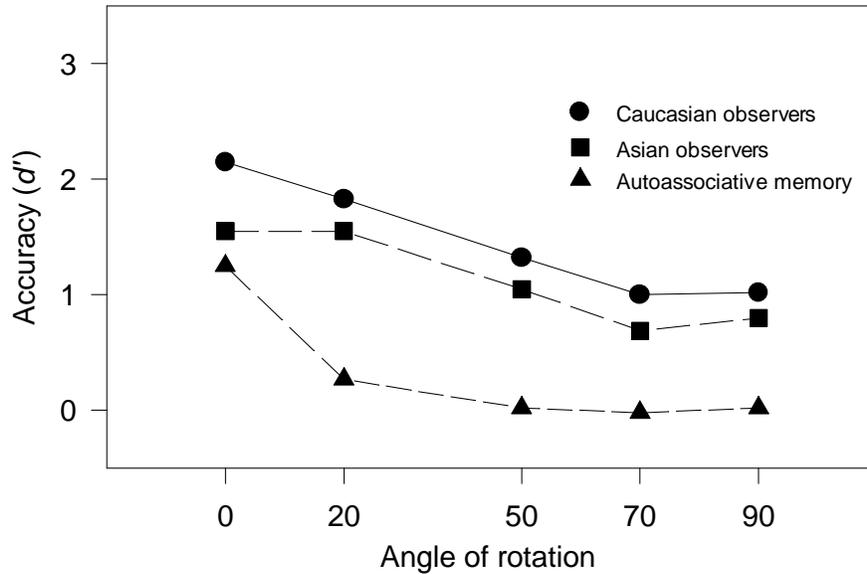


Figure 14. Comparison of the accuracy performance (d') for Caucasian observers (Exp. 1), Asian observers (Exp. 2), and the autoassociative memory. Note that both Asian and Caucasian observers outperformed the computational model.

affected by depth rotation than Caucasian observers recognizing the same faces. In fact, Asian observers even seem to be less affected by a rotation of 20 degrees than are the Caucasian observers. Ellis and Deregowsky reported that Caucasian observers are more sensitive to small changes in orientation when presented with Black faces than when presented with Caucasian faces. However, a more recent study by Ayuk (1990) also failed to replicate the results obtained by Ellis and Deregowsky. Consistent with the results of the present experiment, Ayuk reports that African observers recognizing Caucasian faces are not more affected by a change of orientation than European observers recognizing the same faces (and *vice versa*).

In summary, it would appear, from both the present results and the results reported by Ayuk (1990), that the ability of human observers to extract some invariant information from single views of faces is not linked to their general familiarity with a particular class of faces. Therefore, the poor performance of the autoassociative model, as compared to that of human observers (*cf.* Figure 14), cannot be attributed solely to a difference in general familiarity with Caucasian faces. A correlation analysis showed that, in fact, the squared correlation between the autoassociative memory performance and the Caucasian observers performance was greater than the squared correlation computed between the memory performance and the Asian observers performance (.88 and .70, respectively). This difference is due to the fact that Asian

observers were not affected at all by a 20-degree rotation while both the model and Caucasian observers were. In other words, the performance of the model is closer to that of observers with a higher general familiarity with Caucasian faces than that of observers with a lower familiarity.

General discussion

The experiments and simulations reported in this paper examined and contrasted the ability of human observers and an autoassociative memory to recognize faces presented from a new view angle. Experiment 1 showed that the performance of human observers does not decline as a linear function of the angle of rotation between learning and test. On the contrary, it tends to stay stable up to a 30-degree rotation and drops after this cut-off point to reach an asymptotic value of $d' \approx 1$. Simulations showed that the performance of the autoassociative memory also drops after 30 degrees. However, unlike human observers, the autoassociative memory performs at chance level after this cut-off point ($d' \approx 0$). Experiment 2 showed that the difference in performance between human observers and the autoassociative memory cannot be explained by a difference in general familiarity with the class of faces. Asian observers who are not as familiar with the general category of Caucasian faces were not more affected by a change in orientation between learning and test than Caucasian observers. In fact, the performance of the Caucasian observers was closer to that of the autoassociative memory than the performance of the Asian observers.

An alternative explanation for the difference between human observers and the autoassociative memory performance is that human observers can use two different types of information to perform the recognition task, whereas the autoassociative memory relies on a single type. The data collected in Experiment 1 (d' and C), along with the results of the simulations suggest that, up to 30 degrees, the similarity between the original images and the rotated images is large enough to allow a configural transfer. In this case, both the human observers and the autoassociative memory perform above chance level. After this cut-off point, the similarity between original and rotated faces seems to be too small for a configural transfer to succeed. In this case, only human observers perform above chance level, thus suggesting a different type of transfer. A plausible hypothesis would be that observers transfer some kind of invariant information to recognize faces rotated by more than 30 degrees. Although not as useful as the global configural information presumably used up to 30 degrees, this information would yield performance above chance level. Because the autoassociative memory was not able to capture this invariant information, we can suspect that it is very localized. In agreement with this hypothesis, an examination of the human observers performance obtained for individual faces showed that (a) some faces were always correctly recognized after 90 degrees and others were never correctly recognized and (b) faces correctly recognized tended to have more localized distinctive features (freckles, scars, moles, unusual hair color, *etc.*) than other faces (*cf.* Figure 15).



Figure 15. Examples of faces that were (a) always recognized after a 90-degree rotation (left panels) and (b) never recognized (right panels). Note that the first face has blotchy markings on the right cheek and on the chin that are visible from both the frontal and the profile views. The second face has a very unusual hair color. No such distinctive marks have been detected on the last two faces.

This item analysis suggests that human observers extract some information characteristic of a face that is visible from most viewpoints. During the recognition phase they may use this information to perform what we shall call a transfer by “peculiarity.” When the angle of rotation is smaller than 30 degrees, a face could be recognized using either global configural information or peculiar information—whichever is more convenient. For a rotation larger than 30 degrees, however, only the transfer by peculiarity would be available.

To be efficient, the information used for this second type of transfer has to be specific to a given face. It can range from a very particular hair color or skin texture to a much more localized distinctive feature, such as a scar or a blemish. To decide that a face was present in the learning set, observers need only to recognize the peculiarity of the face without having to recognize the face *per se*. For example, to answer correctly “yes” to the face presented in Figure 16, it suffices to recognize the scar or the lock of white hair. Indeed, this strategy will work only for faces having a peculiarity visible from many different viewpoints (obviously, this will work only until the peculiarity is occluded by the face as it is rotated). This last point is coherent with the decline in performance observed in Experiment 1 and 2 after 30 degrees because, although all faces share configural properties, only a subset of faces contains these peculiarities. Finally, because the transfer by “peculiarity” is, in fact, akin to standard object recognition rather than face recognition, it is not surprising that other-race observers were able to perform this type of transfer.

In summary, our results can be interpreted in terms of a dual-transfer hypothesis

stating that (a) for small rotations, faces would be recognized using either transfer by configuration or by peculiarity depending on faces and on task demands, and (b) for large rotations, only the transfer by peculiarity would be used.

An explicit test of this dual-transfer hypothesis can be found in a recent paper of Valentin, Abdi, and Edelman (in press). The study, described in this paper, investigated the effect of distinctive marks on the recognition of unfamiliar faces across view angles. Subjects were asked to memorize a set of target faces, half of which having distinctive marks. Recognition was assessed by presenting the target faces, either in the same orientation, or after a 90-degree rotation, mixed with an equal number of distractors. The authors found support for the dual-transfer hypothesis for faces learned as frontal views (*i.e.*, marked faces were better recognized, after rotation, than unmarked faces were). However, when a profile view was learned, the presence of marks did not improve recognition performance. The authors attribute this absence of effect to the subjects paying more attention to the shape than to the texture when they see a profile view. Because the marks affect the texture of the face, not paying attention to the texture eliminates the effect of the marks. The authors conclude that the presence of marks, when detected, allows for a recognition by peculiarity.

An additional way of testing the dual-transfer hypothesis could be to show that it is possible to selectively affect performance for small rotations while not affecting performance for large rotations and *vice versa*. For example, we can expect that using low-pass filtered face images will degrade performance for large rotations but not for small ones. Inversely, using high-pass filtered face images should degrade more the performance for small rotations than for large ones.



Figure 16. An example of localized information visible from different view angles.

The remaining question is: How can we model transfer by peculiarity within the framework of an autoassociative memory? Borrowing from Kohonen (1977), we propose the concept of a “novelty filter,” or “peculiarity filter” (*cf.* Figure 17) as a possible answer to this question. As we mentioned previously, the major property of an autoassociative memory is that it behaves as a pattern completion device. When presented with an incomplete image of a face (*e.g.*, mouth area blacked out), an

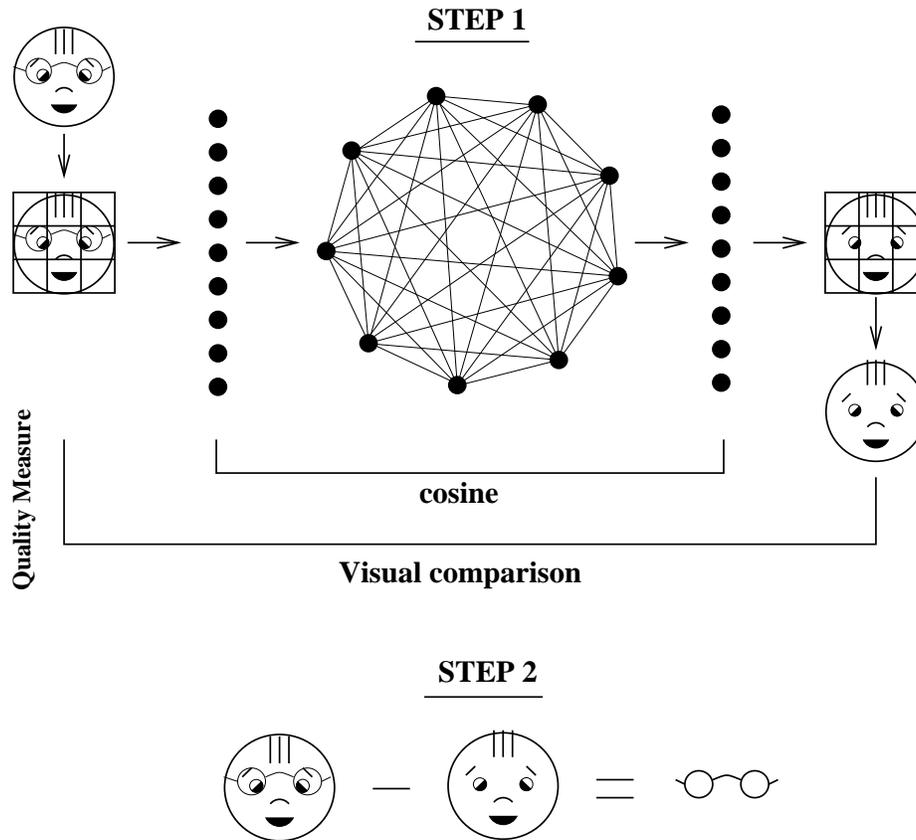


Figure 17. How to implement a novelty filter with an autoassociative memory.

autoassociative memory reconstructs the missing area by replacing the value of each pixel of this area by its expected value. The expected values correspond to a weighted average of the pixel values calculated across the learned faces. Likewise, if we present a face with a pair of glasses (see Figure 17, Step 1) to an autoassociative memory trained to reconstruct glass-less faces, the memory will give back the face without glasses. In other words, the memory will detect what is “peculiar” about the face and discard the peculiarity.

To model a transfer by peculiarity, however, the idea is not to discard the peculiarity, but to isolate it and to use it as a basis for recognition. This can be implemented with an autoassociative memory by using the error of reconstruction of the memory instead of the reconstruction itself. This error of reconstruction is equal to the difference between the original image and the reconstructed image (see Figure 17, step 2). If the face possesses a very distinctive feature, the feature will contribute massively to the error and a simple thresholding technique will make it pop out (see Figure 18, for a more realistic example).

If the detection occurs during the learning phase, the peculiarity will be stored in the memory. If, on the contrary, it happens during the testing phase, a standard

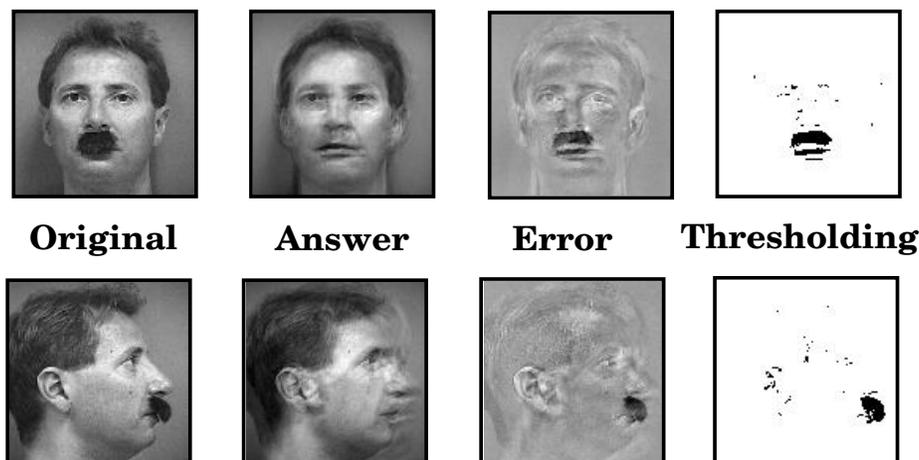


Figure 18. Example of a novelty filter: 1) A mustache is added to a face that was learned by a memory trained to reconstruct clean-shaven faces, 2) the memory reconstructs the face without the mustache (*i.e.*, the pixels corresponding to the mustache area are replaced by the expected value of these pixels), 3) the difference between the original and reconstructed image is computed, and 4) a thresholding technique is used to make the peculiarity pop out.

matching procedure will be used to find out if a similar peculiarity has been stored in the memory. If this is the case, the face will elicit a positive response; if this is not the case, it will elicit a negative response. Thus, the benefit of a novelty filter is twofold. Not only can it lead to the recognition of a face by recognizing the “peculiarity” of this face, but it can also lead to the rejection of a face because none of the learned faces had a similar “peculiarity.”

Although further work will be needed to test systematically the performance of such a novelty filter, (especially with natural features like freckles) this approach constitutes a way of modeling the performance of human observers in a transfer task from single views of faces. If a target face is rotated between learning and test by less than 30 degrees, the global similarity between the learned view of the target and the view presented at test can be used as a basis for the recognition. The cosine between original and reconstructed faces is computed and used as an index of familiarity to decide if the face has been learned or not. In contrast, if the face is rotated by more than 30 degrees, the error produced by the memory will be used as a basis for the recognition. The memory will look for a potential peculiarity and compare this peculiarity with stored peculiarities.

Conclusion

The data reported here show that, although 2D pixel representations are not inherently 3D invariant, they contain enough information to enable recognition across small view angles. If several 2D representations are used simultaneously to represent

a set of faces, a simple linear model is able to interpolate between these views and recognize the faces from a new view angle with an accuracy rate of about 80%. If a single 2D representation is used, the performance of both an autoassociative memory and human observers depends upon the difference between original and rotated views. Specifically, the results of the experiments and computer simulations described in Section 4 suggest that human observers use two different strategies to perform a recognition task from single views of faces. The first strategy, which we call “transfer by configuration,” is probably specific to faces and relies on the global similarity between different views of a face. This strategy, efficient up to a 30-degree rotation, breaks down beyond this point. It can be modeled using a standard autoassociative memory.

The second strategy, which we call “transfer by peculiarity” is not specific to faces and relies on the existence of “peculiarities” or distinctive features visible from most viewpoints. This strategy relies on extracting whatever information (property of the face or of the image) that will allow the observer to perform the task. This strategy could have been called “transfer by whatever” or even “transfer by disfiguration” to emphasize the fact that any difference from an average face visible from different viewpoints would be a very useful cue. Because the success of this type of strategy does not rely simply on the statistical structure of faces, it requires a more sophisticated modeling tool than the simple transfer by configuration. In this case, the autoassociative memory may be used as a kind of novelty filter to isolate the peculiarity of the face. This peculiarity is then treated as an object and recognized using standard object recognition models.

Finally, it is worth noting that our conclusion that two strategies are needed to account for recognition across orientations can be related to some earlier work by Vokey and Read (1992, 1995) and O’Toole, Deffenbacher, Valentin and Abdi (1994), indicating that human observer recognition performance relies on two independent aspects of faces: *memorability*, and *general, or context free familiarity*. As for the dual-transfer hypothesis proposed here, these two aspects of face recognition would be based on different types of perceptual information. Specifically, according to O’Toole *et al.*, faces that are considered very memorable (*i.e.*, the ones that the observers thought would be easy to remember) were characterized by the presence of a small distinctive feature. On the other hand, faces with a low general familiarity level (*i.e.*, the ones that the observers believe they may not have seen around campus) deviate from the set of faces in terms of global face and head shape. If we put together these earlier results and the results presented here, it seems that highly memorable faces should also be highly transferable.

Mathematical foundations

We give in this appendix a short *précis* of the mathematics involved in the PCA model (for more details, see Abdi *et al.*, 1999, in press). We start by capturing images of faces as grey levels. This gives, for each face, a matrix whose elements are

the grey intensity of the pixels of the image. Now the k th face picture is an $M \times N$ matrix denoted \mathbf{Y}_k where M is the number of rows of the image and N its number of columns. This matrix is then transformed into an $I \times 1$ (with $I = M \times N$) vector denoted \mathbf{x}_k with the `vec` operation (*i.e.*, \mathbf{Y}_k is “vectorized” into \mathbf{x}_k):

$$\mathbf{x}_k = \text{vec}\{\mathbf{Y}_k\}. \quad (1)$$

The set of faces to be learned is represented into an $I \times K$ matrix \mathbf{X} in which \mathbf{x}_k is the K column. Each element of \mathbf{x}_k corresponds to the activation level of a neuron like unit of the autoassociative memory. These units are all connected to each other. The values of the connections are given by an $I \times I$ matrix denoted \mathbf{W} (how to find such a matrix is addressed later).

The response of the autoassociator to an input vector is obtained by premultiplying it by the connection matrix. Formally, if \mathbf{x} is an $I \times 1$ vector (which may or may not have been learned), the response denoted $\hat{\mathbf{x}}$ is obtained as

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}. \quad (2)$$

The autoassociator learns or “stores” a pattern by changing the connection matrix \mathbf{W} . This change is obtained by adding, to each synapse, a small quantity which can be a function of the state in which the cells are and of their actual or desired response. There are two main learning rules: Hebb and Widrow-Hoff. Hebbian learning increments the connection between two units proportionally to how similar their activations are for a given pattern. Specifically, storing the k th pattern is obtained as

$$\mathbf{W}_{[t+1]} = \mathbf{W}_{[t]} + \eta \mathbf{x}_k \mathbf{x}_k^T, \quad (3)$$

where η is a small positive number called the *learning constant*. Widrow-Hoff learning first computes the *reconstruction error* as the difference between the expected response (*i.e.*, the input \mathbf{x}_k) and the actual response (*i.e.*, $\hat{\mathbf{x}}_k$). Then, this learning rule increments the connection between two units proportionally to the reconstruction error of the output unit and the activation of the incoming unit. This will make the magnitude of the reconstruction error smaller for subsequent presentations of this pattern. Specifically, storing the k th pattern with Widrow-Hoff learning is obtained as

$$\mathbf{W}_{[t+1]} = \mathbf{W}_{[t]} + \eta (\mathbf{x}_k - \hat{\mathbf{x}}_k) \mathbf{x}_k^T, \quad (4)$$

where η is a small positive number and $(\mathbf{x}_k - \hat{\mathbf{x}}_k)$ is the reconstruction error. Perfect performance for an autoassociator corresponds to perfectly reconstituting stored patterns (but not new patterns). In fact, the better the reconstruction of a pattern, the more likely it is that this pattern was learned (because we want learned patterns to be perfectly reconstructed). To evaluate the quality of the reconstruction, the general strategy is to evaluate the similarity between a pattern and its reconstruction. Similarity is often measured by the cosine between vectors defined as

$$\cos\{\mathbf{x}, \mathbf{y}\} = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad \text{with } \|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}. \quad (5)$$

When the vectors are centered (*i.e.*, when they have zero means) their cosine is equal to their correlation coefficient. The cosine between \mathbf{x} and $\hat{\mathbf{x}}$ is called the *reconstruction cosine*. The larger the reconstruction cosine of a vector, the more evidence there is that it was learned. For a given weight matrix, the vectors with a perfect reconstruction cosine are of particular interest. If they were learned, then they are perfectly recognized. If they were not learned, then they correspond to a maximum false alarm: An effect reminiscent of the prototype effect. It turns out that, for any given weight matrix (obtained from Hebb or Widrow-Hoff learning), it is possible to compute these vectors. They are called *eigenvectors*, or *eigenfaces* (when face images are learned). If \mathbf{u} is an eigenvector of \mathbf{W} , then

$$\lambda \mathbf{u} = \mathbf{W} \mathbf{u} , \quad (6)$$

where λ is a scalar called the *eigenvalue* associated with the eigenvector. In general for autoassociators, the eigenvectors of weight matrices have several remarkable properties. First the eigenvalues are always positive or zero (technically, these matrices are called *positive semidefinite*). Second, eigenvectors with different eigenvalues are orthogonal to each other (*i.e.*, their cosine is zero). Third, the *rank* of a matrix is given by the number of its eigenvectors with non-zero eigenvalues. Finally, a weight matrix \mathbf{W} of rank L , can be decomposed as a weighted sum of its eigenvectors (this is a consequence of being positive semidefinite) as

$$\mathbf{W} = \sum_{\ell}^L \lambda_{\ell} \mathbf{u}_{\ell} \mathbf{u}_{\ell}^{\top} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\top} , \quad (7)$$

where \mathbf{U} is the matrix of eigenvectors and $\mathbf{\Lambda}$ the diagonal matrix of the eigenvalues.

The notion of eigenvectors can also unify Hebbian and Widrow-Hoff learning under a common framework. It can be shown that Widrow-Hoff learning changes only the eigenvalues of the Hebbian learning matrix (the eigenvectors remain unchanged). Specifically (see Abdi, 1994, for a proof), the connection matrix after t learning epochs (an epoch corresponds to learning all the to be learned patterns) is obtained as

$$\mathbf{W}_{[t]} = \mathbf{U} \mathbf{\Phi}_{[t]} \mathbf{U}^{\top} \quad \text{with} \quad \mathbf{\Phi}_{[t]} = \mathbf{I} - (\mathbf{I} - \eta \mathbf{\Lambda})^t , \quad (8)$$

where \mathbf{I} is the identity matrix, \mathbf{U} the matrix of the eigenvectors of the matrix $\mathbf{X} \mathbf{X}^{\top}$, and $\mathbf{\Lambda}$ the matrix of the eigenvalues of $\mathbf{X} \mathbf{X}^{\top}$. When $t = 1$, the connection matrix $\mathbf{W}_{[1]}$ is the Hebbian connection matrix. This shows that the matrices obtained from these two learning rules differ only by the value of an exponent.

Learned vectors can be built as a sum of the eigenvectors of the connection matrix. Specifically, if \mathbf{x}_k was learned then

$$\mathbf{x}_k = \sum_{\ell}^L \gamma_{\ell,k} \mathbf{u}_{\ell} \quad \text{with} \quad \gamma_{\ell,k} = \mathbf{u}_{\ell}^{\top} \mathbf{x}_k . \quad (9)$$

References

- Abdi, H. (1987). *Introduction au Traitement des Données Expérimentales*. Grenoble: Presses Universitaires de Grenoble.
- Abdi, H. (1988). Generalized approaches for connectionist autoassociative memories: Interpretation, implication, and illustration for face processing. In Demongeot, J., *Artificial Intelligence and Cognitive Sciences*. (pp. 151–164). Manchester: Manchester University Press.
- Abdi, H. (1994a). *Les réseaux de neurones*. Grenoble: Presses Universitaires de Grenoble.
- Abdi, H. (1994b). A neural network primer. *Journal of Biological Systems*, **2**, 247–281.
- Abdi, H., Valentin, D. and Edelman, B. (1999). *Neural networks*. Thousand Oaks, CA: Sage.
- Abdi, H., Valentin, D. and Edelman, B. (in press). *Neural networks for cognition*. Sunderland, MA: Sinauer.
- Abdi, H., Valentin, D., Edelman, B. and O’Toole, A. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception*, **24**, 539–562.
- Ayuk, R. (1990). Cross-racial identification of transformed, untransformed, and mixed-race faces. *International Journal of Psychology*, **25**, 509–527.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, **94**, 115–145.
- Bruce, V. (1982). Changing faces: Visual and non-visual coding process in face recognition. *British Journal of Psychology*, **73**, 105–116.
- Edelman, B., Valentin, D., and Abdi, H. (1998). Sex classification of faces by human subjects and a neural network. *Journal of Biological Systems*, **6**, 241–263.
- Ellis, H. and Deregowski, J. (1981). Within-race and between-race recognition of transformed and untransformed faces. *American Journal of Psychology*, **94**, 27–35.
- Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Hasselmo, M., Rolls, E., Baylis, G. and Nalwa, V. (1989). Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, **79**, 417–429.
- Kohonen, T. (1977). *Associative memory: A system theoretic approach*. Berlin: Springer.
- Macmillan, N. and Creelman, C. (1991). *Detection theory: A user’s guide*. Cambridge: Cambridge University Press.

- Marr, D. and Nishihara, H. (1978). Representation and recognition of the spatial organization of three dimensional shape. *Proceedings of the Royal Society of London B*, **200**, 269–294.
- McNicol D. (1972). *A primer in signal detection theory*. London: Allen & Unwin.
- Millward, R. and O’Toole, A. (1986). Recognition memory transfer between spatial-frequency analyzed faces. In H. Ellis, M. Jeeves, F. Newcombe, and A. Young, *Aspects of face processing*. (pp. 34–44). Dordrecht: Nijhoff.
- O’Toole, A., Abdi, H., Deffenbacher, K. and Valentin, D. (1993). A low dimensional representation of faces in the higher dimensions of the space. *Journal of the Optical Society of America A*, **10**, 405–411.
- O’Toole, A., Abdi, H., Deffenbacher, K. and Valentin, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine, *Cognitive and computational aspects of face recognition*. London: Routledge.
- O’Toole, A., Deffenbacher, K., Abdi, H. and Bartlett, J. (1991). Simulating the other race effect as a problem in perceptual learning. *Connection Sciences*, **3**, 163–178.
- O’Toole, A.J., Deffenbacher, K.A., Valentin, D. and Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory and Cognition*, **22**, 208–224.
- Perrett, D., Mistin, A., Potter, D., Smith, P., Head, A., Chitty, A., Broennimann, R. Milner, A. and Ellis, M. (1986). Functional organization of visual neurons processing face identity. In H. Ellis, M. Jeeves, F. Newcombe and A. Young, *Aspect of face processing*, pages 187–198. Dordrecht: Nijhoff.
- Perrett, D., Rolls, E. and Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, **47**, 329–342.
- Rosch, E. (1978). Principles of categorization. In E. Rosch and B. Lloyd, *Cognition and Categorization*, pages 27–48. Hillsdale: Erlbaum.
- Shepard, R. and Cooper, L. (1982). *Mental images and their transformations*. Cambridge: The MIT Press.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neurosciences*, **3**, 71–86.
- Valentin, D. and Abdi, H. (1996). Can a linear autoassociator recognize faces from new orientations? *Journal of the Optical Society of America A*, **13**, 717–724.
- Valentin, D., Abdi, H. and Edelman, B. (1997). What represents a face: a computational approach for the integration of physiological and psychological data. *Perception*, **26**, 1271–1288.
- Valentin, D., Abdi, H. and Edelman, B. (in press). From rotation to disfiguration: Testing a dual-strategy model for recognition of faces across view angles. *Perception*, **28**.

- Valentin, D., Abdi, H., Edelman, B. and O'Toole, A. (1997). Principal component and neural network analyses of face images: What can be generalized in gender classification? *Journal of Mathematical Psychology*, **41**, 398–413.
- Vokey, J.R. and Read, J.D. (1992). Familiarity, memorability and the effect of typicality on the recognition of faces. *Memory and Cognition*, **20**, 291–392.
- Vokey, J.R. and Read, J.D. (1995). Memorability, familiarity and categorical structure in the recognition of faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition*. (pp. 113–137). London: Routledge.