Perception, 1997, volume 26, pages ***–***

What Represents a Face: A Computational Approach for the Integration of Physiological and Psychological Data

Dominique Valentin[*†], **Hervé Abdi**[*†] **Betty Edelman** [*]

[*] **School of Human Development, The University of Texas at Dallas, Richardson, TX 75083-0688, USA.**
[†] **Université de Bourgogne, 21004 Dijon Cedex, France.**

Abstract. Empirical studies of face recognition suggest that faces might be stored in memory using a few canonical representations. The nature of these canonical representations is however unclear. Although psychological data show a 3/4 view advantage, physiological studies suggest profile and frontal views are stored in memory. In this paper we propose a computational approach to reconcile these findings. The patterns of results obtained when different views, or combinations of views, are used as the internal representation of a two-stage identification network consisting of an autoassociative memory followed by an RBF network are compared. Results show that 1) a frontal and a profile view are sufficient to reach the optimal network performance; 2) all the different representations produce a 3/4 view advantage, similar to that generally described for human subjects. These results indicate that although 3/4 views yield better recognition than other views, they need not be stored in memory to show this advantage.

1. Introduction

How do we recognize familiar faces from a variety of viewpoints? This is a fundamental problem from both psychological and computational perspectives. As a face is subjected to rotation in depth, its retinal projections change drastically. Yet human observers seem to have very little difficulty in recognizing familiar faces from most view points. Several questions arise concerning the memory representation that supports this expertise. For example, how can we recognize that each of the images presented in Figure 1 represents the same face viewed from different angles? Do we store some canonical or prototypical views of the face? Or do we instead store a whole set of images or descriptions to cover every contingency?

The main purpose of the work presented here is to explore, using numerical simulations, the usefulness of different types of memory representations for generalizing across view orientations. This paper is organized as follows. First, we present some psychological and neurophysiological studies relevant to the problem of facial representations. These studies converge to the idea that faces are stored in memory using a limited set of 2D views of faces, but disagree on the nature of

FIGURE 1. Illustration of a face rotated in depth. how do we know that each of the images represents the same face?

these views. Second, we report a human-subject experiment and a series of simulations designed to explore further the discrepancies observed in the literature.

The goal of the human-subject experiment is to replicate, for our face database, some previous findings reported in the psychological literature. The goal of the series of simulations is to compare the recognition/identification performance obtained when different views or combinations of views are used as the internal representation of a computational model. A two-stage neural network, consisting of an autoassociative memory followed by a radial basis function (RBF) network, is trained to "identify" a set of faces presented from different view angles. The ability of the model to generalize to new views of the faces is then tested. The patterns of results yielded by different internal representations are contrasted and compared with the human-subject data.

2. Previous work

Very little research has been directed to exploring specifically or explicitely the kind of facial representation human observers store in long-term memory. However, although not always designed to investigate this specific problem, diverse studies have provided some insight into the nature and properties of the representation that might be developed for familiar and unfamiliar faces. These studies have been conducted simultaneously in the fields of psychology and neurophysiology with the goal of examining the effect of depth rotation on face recognition.

2.1. *Psychological studies.* A first illustration of the ability of human observers to handle depth rotation was provided by an experiment performed by Patterson and Baddeley (1977) using unfamiliar faces. They demonstrated that a change in both orientation and expression (from full-face unsmiling to 3/4 smiling) between learning and test did not affect subjects' performance significantly. Subjects were able to identify transformed faces at a level equivalent to that obtained with untransformed faces. However, if the change involved a greater depth rotation (e.g. from full-face to profile), subjects were less accurate in identifying the faces. Using a recognition task, Davies, Ellis and Shepherd

(1978) also found that altering the orientation of faces from frontal to 3/4 views (and *vice versa*) between learning and test did not affect recognition accuracy.

These studies suggest that single views of faces contain enough invariant information to allow for recognition and identification of the faces over moderate changes in view angles (up to 45 degrees) between learning and test. However, these results have not always been replicated. For example, Baddeley and Woodhead (1981) found a decrement in recognition accuracy when faces were changed from a frontal view to a 3/4 view between learning and test and *vice versa*. Using a similar approach with both familiar and unfamiliar faces, Bruce (1982) reported also that, for unfamiliar faces, changing the view from full-face unsmiling to 3/4 smiling between learning and testing reduced the accuracy and increased the latency of the subjects' responses. For familiar faces, an effect of view transformation appeared only on the response latency.

To test the hypothesis that faces could be recognized more easily in some orientations than in others, Krouse (1981) presented a group of subjects with frontal or 3/4 views of a series of unfamiliar faces. The subjects were then asked to recognize the faces presented, either in the same or in a different orientation. In addition to the classical effect of view change between study and test, Krouse found a significant effect of *study view*. Three-quarter views at presentation led to better performance at test than frontal views. Using a similar paradigm, Logie, Baddeley and Woodhead (1987) found that an initial study of a 3/4 view led to better recognition performance than either a frontal or a profile view. This advantage for the 3/4 view was also found with babies in an earlier study by Fagan (1979) using a habituation paradigm. Babies presented with adult faces at different orientations showed better recognition performance with 3/4 views than with frontal or profile views of the faces.

Bruce, Valentine, and Baddeley (1987) investigated whether the 3/4 view advantage could be extended from a recognition task performed with unfamiliar faces to a speeded recognition task using familiar faces. In a first experiment, they tested whether highly familiar faces could be categorized as familiar more readily when presented in 3/4 rather than in frontal views. Results showed no evidence of a 3/4 view advantage for accepting familiar or for rejecting unfamiliar faces. In a second experiment, subjects were presented with pairs of faces and asked to indicate whether the faces were of the same person or of different persons. In both same and different trials, the two faces had different facial expressions (smiling and neutral). They observed a 3/4 view advantage for unfamiliar faces on positive trials: Two 3/4 views were matched more quickly than were two frontal views. Matchings of profiles were slowest of all. This 3/4 view advantage was not observed on negative trials or for familiar faces. Similar results have been since reported by Bruyer and Galvez (1989).

In summary, the picture that emerges from psychological studies is a rather confusing one. Some studies indicate that a change in orientation from frontal to 3/4 view between learning and test does not affect recognition accuracy (Davies et al. 1978; Patterson & Baddeley, 1977). Other studies found that such a change did affect the recognition performance for unfamiliar faces (Baddeley & Woodhead, 1981; Bruce, 1982) but not for familiar faces (Bruce, 1982). A potential explanation for this diversity of results is that depth rotation does not affect all faces in a similar way (i.e. there is an item effect for faces). However, despite these divergences, the studies reported in this section suggest a shift in perceptual representation from unfamiliar faces to familiar faces. Whereas recognition/identification performance for familiar faces tends to be insensitive to depth rotation, recognition/identification performance for unfamiliar faces tends to decrease as faces are rotated in depth. In terms of internal representation, these results make the hypothesis of a 3D invariant representation questionable. Moreover, the fact that for unfamiliar faces 3/4 views lead to better recognition performance than either frontal or profile views suggests a view-dependent representation as a better candidate than a 3D invariant representation. This 3/4 view advantage can be interpreted

as an indication that 3/4 views are "canonical" views of faces, as defined by Palmer, Rosch and Chase (1981) for object recognition. According to these authors, certain views of an object are judged *better views* of the object than certain other views. These "canonical" views maximize the amount of salient information for the object. They lead to faster identification and are spontaneously reported by human observers. Bruyer and Galvez (1989), for example, suggest that 3/4 views of faces could constitute the "structural code" proposed by Bruce and Young (1986) and Fagan (1979) notes that 3/4 views are more salient than either frontal or profile views. However, as noted by Bruce (1988), no real support has been found, yet, for such an interpretation (see also Bruce et al., 1987, for a discussion).

2.2. *Neurophysiological data.* The neurophysiological studies relevant to face representation consist of single cell recordings in the temporal cortex of monkeys presented with either monkey or human faces from different orientations. These studies have been applied both to the problem of depth rotation (i.e. how does depth rotation affect the responses of single cells?) and to the problem of the existence of canonical views (i.e. are there some cells preferentially tuned to particular views of faces?). Although the results of these studies cannot be used as direct evidence for human subjects, they can help us to understand some process underlying face recognition, and lead to new hypotheses of face representation.

For example, Perrett, Rolls, and Caan (1982, see also Desimone, Albright, Gross & Bruce, 1984; Perrett et al., 1985; Perrett et al., 1986; Perrett, Mistlin & Chitty, 1987) 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. Among the face-specific cells, some cells, or groups of cells, responded to specific faces across different viewing orientations. Other cells responded to many different faces but were sensitive to depth rotations. Rotating the faces from frontal to profile views reduced or eliminated the response of 60% of these cells. Even rotations as small as 10 or 20 degrees produced a substantial reduction of the responses.

Some of these view-specific cells were tuned to frontal, others to profile views, and some others to the back of the head. Interestingly, they found no cells specifically tuned to views intermediate between full-face and profile. Perrett et al. (1986) interpreted this finding as an indication that " ... the recognition of each individual known to an observer proceeds by an analysis of a small set of prototypical views of that individual" (p.191). They theorized that intermediate views are recognized by interpolating between these prototypical or canonical views. For example, they mentioned that even in the absence of 3/4 view-specific cells, 3/4 views might generate the same amount of total responses as frontal or profile views by activating both full-face *and* profile specific cells to half the rate produce by the preferred "canonical" views.

Of course the fact that they did not find cells preferentially tuned to intermediate views of faces cannot be taken as a proof for the nonexistence of such cells. Indeed, later studies discovered a few cells specifically tuned to other views than the canonical views mentioned by the early studies, thus casting a doubt on the original claim of preferential coding of views of faces (Hasselmo, Rolls, Baylis, & Nalwa, 1989; Perrett, Mistlin & Harries, 1989; Perrett et al., 1991). Yet, Perrett, Oram, Hietanen & Benson, 1994, report that "recent quantitative and extensive studies have, however, confirmed the notion of preferential coding of particular views. Although cells are tuned to a whole range of views in the horizontal plane there is statistical preference for the face and profile" (pp. 50–51.) In terms of facial representations, this finding suggests that frontal and profile views might be stored in memory and that recognition from other views could be done mostly by interpolating between these two canonical views.

2.3. *Integrating psychological and physiological data.* Both psychological and neurophysiological data suggest that faces are stored in memory using view-dependent representations. However, the

nature of the views stored in memory is not clear. Psychological data suggest the existence of one canonical view: 3/4 view. Physiological data suggest the existence of two canonical views: frontal and profile.

This apparent contradiction between the psychological and neurophysiological data has already been mentioned by Bruce (1988) who indicated as a possible resolution that "Paradoxically, a system which separately represented full-face and profile views could show an advantage for 3/4 views if these were within range of both sets of specialist detectors" (p. 89). However, the estimates of tuning of characteristic cells provided by Perrett et al. (1991) do not give a straight forward support for Bruce's explanation. According to Perrett et al., "for most (characteristic) cells, 45-90 degrees rotation of the head reduced the magnitude of response to half that of the optimal view" (p. 160). At best, with such tunings, a simple additive model would predict that the 3/4 views would be as well recognized as the full face or profile views but not *better*. Moreover, the physiological data alone could not lead *a priori* to the prediction of a 3/4 view advantage.

This contradiction might come from the fact that we are comparing results that are not directly comparable. Finding a way to compare psychological and neurophysiological data is not trivial. First, single cell recordings of face recognition have not been reported for human subjects. Second, no transfer experiments exploring the 3/4 view advantage have been reported for monkeys. Therefore it seems that, isolated, traditional laboratory experiments are unlikely to be helpful in resolving this issue of representation. An alternative to the difficult comparison of psychological and neurophysiological data can be provided by computational simulations. In addition to simulating behavioral data, computational models permit the manipulation of internal representations as well as the simulation of single cell recording data.

The goal of the simulation we present in this paper is to integrate psychological and neurophysiological data. However, because of the item effect often encountered with faces, it is important, first, to show that the particular set of faces used as input to the computational model yields behavior similar to that reported in previous work.

3. **Human experiment**

The purpose of this experiment is to evaluate the ability of human observers to recognize familiar and unfamiliar faces across orientation changes as well as to assess the presence of a 3/4 view advantage for the faces in our database. Most of the previous studies used only a small range of transformations of the faces (45 degrees), which limits the extent of the claim that face recognition is resistant to orientation transformation. In the present recognition experiment subjects are asked to memorize a set of faces presented from a single point of view (either full-face, 3/4 view, or profile). Their memory is then tested by presenting the same faces in one of the three viewpoints mixed with an equal number of distractor faces. The overall range of transformation in this experiment is, thus, 90 degrees. In addition, to verify the fact that familiar and unfamiliar faces are differently affected by a change in orientation between learning and test, two familiarity conditions are used—*familiar* (i.e. subjects knew the faces from somewhere else) and *unfamiliar* (i.e. subjects did not know the faces before the experiment.)

3.1. *Methodology.*

3.1.1. *Observers.* Because the same faces were used in the two familiarity conditions, two different types of observers participated in the experiment. For the unfamiliar condition, 24 undergraduates from the University of Texas at Dallas 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. Only the data of subjects not familiar with the faces were analyzed. For the

0 degrees		45 degrees		90 degrees	
learning	testing	learning	testing	learning	testing
full-face	full-face	full-face	3/4	full-face	profile
3/4	3/4	3/4	full-face	profile	full-face
profile	profile	3/4	profile		
		profile	3/4		

TABLE 1. Patterns of rotation used in the human subject experiment

familiar condition, 24 volunteer graduate students, staff and faculty members of the School of Human Development (UTD) familiar with the faces participated in the experiment.

3.1.2. *Stimuli.* Forty female volunteers were photographed to create a face database. Each face in the database was represented by 20 views including: 1) one series of 10 views sampling the rotation of the head from full-face to right profile with about 10-degree steps, and 2) two series of five views, both sampling the rotation of the head from full-face to right profile with about 20-degree steps. Before the experiment, graduate students, staff and faculty members of the School of Human Development were asked to fill out a brief questionnaire to assess their familiarity with the persons in the database (e.g. from where they knew the person, how long they knew her, how frequently they usually see her). Using the results of this survey, the 30 faces judged most familiar were selected from the database to be used as experimental stimuli. The ten remaining faces were used as *fillers* during the learning session. Performance on these faces was ignored in the analyses of the results. Six pictures of each face served as stimuli (3 view angles $\times$ 2 poses) so that different pictures of the target faces were used for learning and testing in each angle of rotation condition.

3.1.3. *Experimental design.* Forty-eight observers were tested on a standard yes-no recognition task. Two independent variables were manipulated: *familiarity* of the subjects with the faces (unfamiliar *versus* familiar) and *degrees of rotation* between learning and test (0, 45, and 90 degrees). The patterns of transformations used to obtain the different rotation conditions are described in Table 1.

A counterbalancing Latin Square procedure was used to ensure that every face appeared equally often as target and distractor and in each transformation condition. For both learning and testing lists, the order of presentation of the faces was randomized and a different order used for each subject.

3.1.4. *Procedure.* The experiment consisted of two sessions, learning and testing, separated by a 10-minute break. During the learning phase subjects were shown 25 faces (15 targets and 10 fillers), each presented on a computer screen for 4 seconds, with a 4-second interstimulus interval. Approximately one third of the faces were presented from a frontal view, one third from a 3/4 view and the last third from a profile view. Subjects 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 in the test would not necessarily be taken from the same view angle as in the original presentation. During the testing phase, subjects were shown a second series of 40 faces, the 15 targets presented during the learning phase, 15 distractors, and the 10 fillers. For one third of the subjects in the unfamiliar condition and one third of the subjects in the familiar condition, the view orientation of the target faces was the same as in the learning session. For the second third of the subjects in both familiarity conditions, the target faces were rotated in depth by 45 degrees. For the remaining subjects, the target faces were rotated in depth by 90 degrees. For each face, subjects were asked to decide whether they had seen the face during the learning session. They 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 subjects indicated their answer by pressing one of the mouse buttons.

3.2. *Results.* Results were analyzed using signal detection methodology. Each subject contributed a d' calculated on the basis of 15 scores. Hit rates of 100 percent and false-alarm rates of 0 percent were converted to $1 - 1/2N = .97$ and $1/2N = .03$ respectively (cf. Macmillan & Creelman, 1991), thus leading to a maximum value of d' equal to 3.76. Separate analyses of variance (ANOVA) were carried out for estimating: 1) the effect of depth rotation on accuracy performance, and 2) the effect of learning and testing views on accuracy performance.

3.2.1. *Depth rotation.* The mean d' values are shown in Figure 2. A 2×3 between-subjects ANOVA (familiarity $\times$ angle of rotation between learning and test) reveals a highly significant effect of familiarity, $F(1, 42) = 27.08$, $MS_e = .52$, $p < .0001$, and a significant effect of degrees of rotation, $F(2, 42) = 6.11$, $MS_e = .52$, $p < .01$. No significant interaction was observed between familiarity and depth rotation. However, a sub-design analysis of the effect of depth rotation conditional on familiarity shows that the effect of rotation is significant only in the unfamiliar condition, $F(2, 42) = 5.75$, $MS_e = .52$, $p < .01$. Further, 91% of the effect of depth rotation within the unfamiliar condition is due to the difference between 0 and 45 degrees.

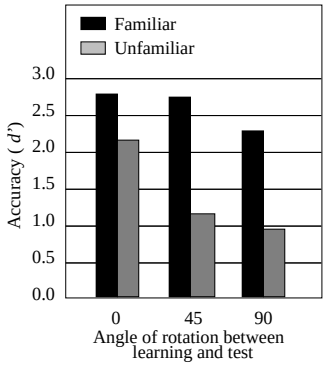

FIGURE 2. Average d' averaged across subjects as a function of familiarity of the subjects and degrees of rotation between learning and test.

3.2.2. *Type of views.* Figure 3 shows the d' values obtained for familiar and unfamiliar subjects as a function of the views presented during learning and testing. A $3 \times 3 \times 2$ between-subject ANOVA with *learning view*, *testing view*, and *familiarity* as independent variables and d' as a dependent variable reveals a main effect of familiarity $F(1, 126) = 33.94$, $MS_e = 1.06$, $p < .0001$, a main effect of type of view at test, $F(2, 126) = 5.13$, $MS_e = 1.06$, $p < .01$, and an interaction between type of view at learning and type of view at test, $F(4, 126) = 9.60$, $MS_e = 1.06$, $p < .05$. A contrast analysis indicates that subjects are more accurate when tested with a 3/4 view than with any of the two other views, independently of the view presented during learning, $F(1, 126) = 4.66$, $MS_e = 1.06$, $p < .05$. No difference was observed between profile and frontal views, $F < 1$. A sub-design analysis by familiarity condition indicates that the 3/4 view advantage is significant only for the unfamiliar condition $F(1, 126) = 6.89$, $MS_e = 1.06$, $p < .01$.

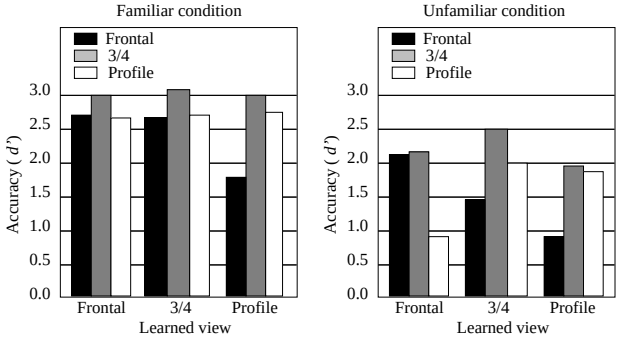

FIGURE 3. Average d' as a function of the familiarity of the subjects with the faces and the view presented during learning and test.

3.3. *Discussion.* The main findings of this experiment are the following. First, as expected, and in agreement with Bruce (1982), changing the orientation of faces between learning and test affects the recognition accuracy of subjects in the unfamiliar condition, but not in the familiar condition. More surprisingly and contrary to previous evidence, for the particular set of faces used here, the effect of degrees of rotation is mainly due to the difference observed between 0 and 45 degrees. Previous work suggested that subjects' performance should stay relatively stable up to 45 degrees of rotation and dramatically decrease beyond this point.

Second, also in agreement with previous work (Logie et al., 1987; Krouse, 1981), 3/4 views of faces yield better accuracy performance than either profile or frontal views. However, a difference should be noted, between the present result and the results reported in the literature. Previous work indicated that a 3/4 view presented during learning led to better performance than a frontal or a profile view. Here the 3/4 view advantage was observed when a 3/4 view was presented *at test*, independent of the view presented during learning. This latter result, however, is not in disagreement with the canonical view hypothesis proposed to account for the 3/4 view advantage described in the literature. Moreover, the fact that, in the present experiment a clear advantage was observed in the 3/4 - 3/4 transfer condition, as compared to the frontal-frontal and profile-profile transfer conditions, provides some additional support for the idea that 3/4 views constitute an optimal view for face recognition. The problem with interpreting this result as indicating that 3/4 views are stored in memory and play the role of canonical views for faces is addressed in the following series of simulations.

4. Computational model

The computational model used to investigate the apparent paradox arising from the 3/4 view advantage of human subjects and single cell recording data consists of an autoassociative memory and an RBF network. The autoassociative memory (see appendix A) is used to preprocess pixel images, yielding a compressed representation of faces. This compressed representation is then input into the RBF network.

4.1. *Autoassociative memory preprocessing.* A first problem in simulating face recognition or identification is to find a way of coding the perceptual information in faces. Previous work showed that autoassociative memories operating on pixel arrays, or more generally the principal component analysis (PCA) approach, provide an efficient solution to that problem (Abdi, 1988; Abdi, Valentin,

Edelman & O'Toole, 1995; O'Toole, Abdi, Deffenbacher, Valentin, 1993; Sirovich and Kirby, 1987; Turk and Pentland, 1991). In this framework, face images are represented by their projections onto a set of statistically derived dimensions. In neural network terms, the dimensions are the eigenvectors of the between unit connection weight matrix (see appendix A). In statistical terms the dimensions are the principal components of the pixel-by-pixel face cross-product matrix. The eigenvectors, or principal components, constitute an orthogonal basis (or eigenspace) for representing the faces. A given face can be either perfectly represented by using the complete eigenspace, or approximated by using a subset of the eigenspace.

Most of the earlier models using this type of approach represented faces using single frontal (or nearly frontal) 2D representations of faces. As a consequence, their performance across large changes in orientation was rather poor. Recent studies (Pentland, Moghaddam, & Starner, 1994; Valentin & Abdi, 1996), however, showed that this limitation can be overcome by using multiple views of the faces. In their study, Pentland et al. used 2D images taken from different view angles, sampling the rotation of the head from left profile to right profile. They created separate covariance matrices for each view angle and decomposed them into their eigenvectors and eigenvalues. This procedure gives rise to a series of eigenspaces for representing the faces, which the authors call view-based eigenspaces. When a face is presented as input, its orientation is first evaluated by computing its distance from all the eigenspaces. Then the face is projected onto the closest eigenspace and identified using a nearest neighbor algorithm. The problem with this approach is that it assumes two separate mechanisms sequentially ordered: determination of the pose followed by identification of the face. There is no psychological evidence for such a dual mechanism.

Valentin and Abdi (1996) proposed a somewhat different approach in which a single eigenspace is computed from multiple views of faces. They showed that this eigenspace provides enough information to both estimate the orientation of the faces and identify them. Their approach has the advantage of avoiding the assumption of separate sequential mechanisms while implementing a natural dissociation between orientation and identity information. The projections of a face onto the first 20 eigenvectors allow for the determination of its orientation, and the projections onto the remaining eigenvectors allow for the identification of the face. A similar approach was used here to represent the faces before using them as input to the RBF network

4.2. *RBF network.* An RBF network is a 3-layer network in which the hidden layer performs a nonlinear mapping of the input layer onto the output layer. Intuitively, the inner workings of an RBF network can be separated into two phases as illustrated by Figure 4. During the first phase (recoding phase) the input patterns are recoded in terms of their distances from prototypes (or centers of the RBF network.) During the second phase (learning phase), the recoded patterns are associated with the expected outputs using a standard heteroassociator (see appendix B for more detail.)

- *Recoding phase.* The recoding of the input is performed by the hidden units of the RBF network. Each hidden unit computes a Gaussian transformation of the distance from the input to its center. From a psychological point of view, the centers of the hidden units can be regarded as some kind of prototypes and the distance from the input to the centers as an indication of the similarity between the input and the prototypes. The activity of a hidden unit depends on the similarity between the exemplar presented as input and its center: When the exemplar matches exactly the center or prototype, the activity of the unit is at its maximum and it decreases as an exponential function of the squared distance between the input and the center. At the end of the recoding phase, the input patterns are represented by I-dimensional vectors (with I representing the number of hidden units) in which a given element represents the activity of a given hidden unit.

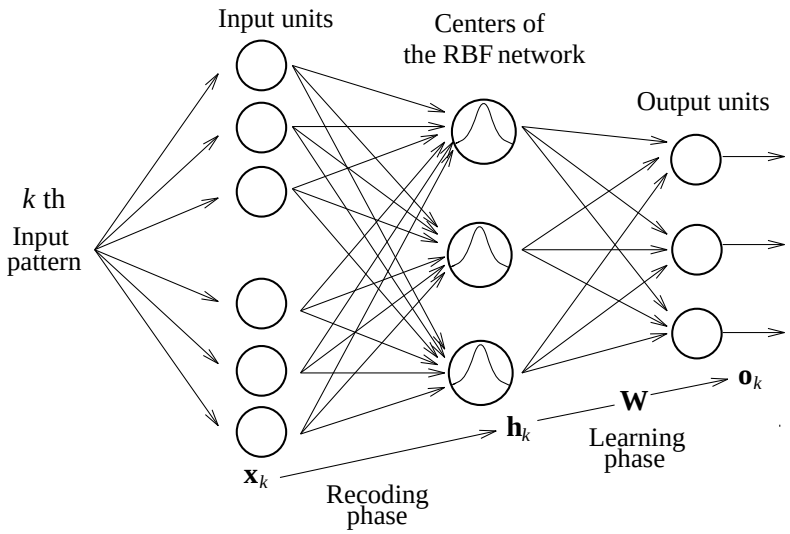

FIGURE 4. Illustration of an RBF network.

- *Learning phase.* During the second step, a linear heteroassociation is performed between the output of the hidden units (or recoded input) and the expected output. In brief, the output units compute a weighted sum of their input (hidden unit outputs.) The weights are the strength of the connections between hidden and output units. Learning is achieved by modifying the connection weights so as to minimize the difference between the expected output and the actual output. The optimum values for the weights can be obtained by using a least-squares approximation.

A useful property of RBF networks for the problem of face representation is that their hidden layers act as internal representations that can be manipulated. By using specific views, or combinations of views, as centers of an RBF network, we can test the usefulness of these views to represent and identify faces. Moreover, when particular views of faces are used as centers, the hidden-units become somewhat similar to the view-specific cells reported in the neurophysiological literature. Specifically, as is the case for view-specific cells, the hidden units can be regarded as preferentially tuned to specific views of faces: Their activity is a function of the similarity between the view presented as input and their center (or preferred view.) This property of RBF networks is used here to compare the patterns of results yielded by different types of internal representations. Five different types of "representations" are used as the centers of a series of RBF networks. The choice of the centers was dictated by different theoretical hypotheses:

1. To test the hypothesis that faces are represented in memory by several 2D views of the faces corresponding to the familiar orientations of the face (exemplar model), we used as many centers as views presented as input for each face.
2. In an attempt to create a view-independent representation of the faces, we used one center for each face: the average of the face across views.
3. To test the different canonical view hypotheses, we used either:
   - One center for each face: a 3/4 view
   - Two centers for each face: a frontal and a profile view
   - Three centers for each face: a frontal, a 3/4, and a profile view.

Each RBF network was trained to identify a set of faces presented from a different number of view angles and tested for its ability to recognize the faces from a new orientation.

5. Simulations

A series of simulations applied the computational model described in the previous section to the task of identifying faces from new orientations.

5.1. *Method.*

5.1.1. *Stimuli.* The 40 faces used in the human subject experiment were used as stimuli. Each face was represented by either 10 views sampling the rotation of the head from full-face to profile in about 10-degree steps (10-view condition), or five views sampling the rotation of the head in about 20-degree steps (5-view condition.)

5.1.2. *Experimental design.* The task simulated was an identification task from new orientations. Two variables were manipulated: The *number of views* presented during learning (four or nine views) and the *type of centers* of the hidden units of the RBF networks (all views; average view; frontal and profile views; frontal, 3/4, and profile view; and 3/4 view only.) For each simulation, the number of identification errors was recorded.

5.1.3. *Procedure.* A two-stage network, consisting of an autoassociative memory followed by an RBF network, was used to identify all the faces in the database using a sample of the available views. The procedure included three steps: A compression step, a learning step, and a testing step. These three steps are illustrated in Figure 5, and are described below:

- *Compression step.* All the views of the 40 faces (five in the 5-view condition, and ten in the 10-view condition) were stored in an autoassociative memory and the memory decomposed into its eigenvectors and eigenvalues (see appendix A). Each view of the faces was then represented by a 50 dimensional vector corresponding to its projections (or weights) onto the first 50 eigenvectors. Figure 6 illustrates the reconstruction of a face with the first 50 eigenvectors. The squared coefficient of correlation between the frontal view (top panels) and its reconstruction is .82. The squared coefficient of correlation between the profile view and its reconstruction is .81. Clearly these reconstructions contain information relative to both the orientation and the identity of the faces[1].
- *Learning step.* All but one view of the 40 faces (4 in the 5-view condition and 9 in the 10-view condition) were used as input to an RBF network made of 50 input units, from 40 to 360 hidden units depending on the learning condition and the type of centers, and 40 output units. In the 5-view condition, four views *per* face were chosen from the set of five possible views (0, 20, 45, 70, and 90 degrees from full-face) to be used as a training set. In the 10-view condition, nine views *per* face were chosen from the set of 10 possible views (0, 10, 20, 30, 40, 50, 60, 70, 80, and 90 degrees from full-face) to be used as a training set. In both conditions, the training views were randomly chosen, under the constraint that each view angle appear an approximately equal number of times in the learning list. The remaining view was used as a testing view.

  The network was trained to produce a 1 in the output unit corresponding to the face presented at the input layer and a 0 in every other unit. For example, if as illustrated in

[1]The ghostly appearance of the reconstruction is due to the fact that only 50 eigenvectors out of 200 have been used. The appearance of the image can be improved by increasing the number of eigenvectors. However, doing so does not improve the identification performance of the RBF network, thus showing that the visual appearance of the reconstructions and the usefulness of the information they convey are two separate issues.

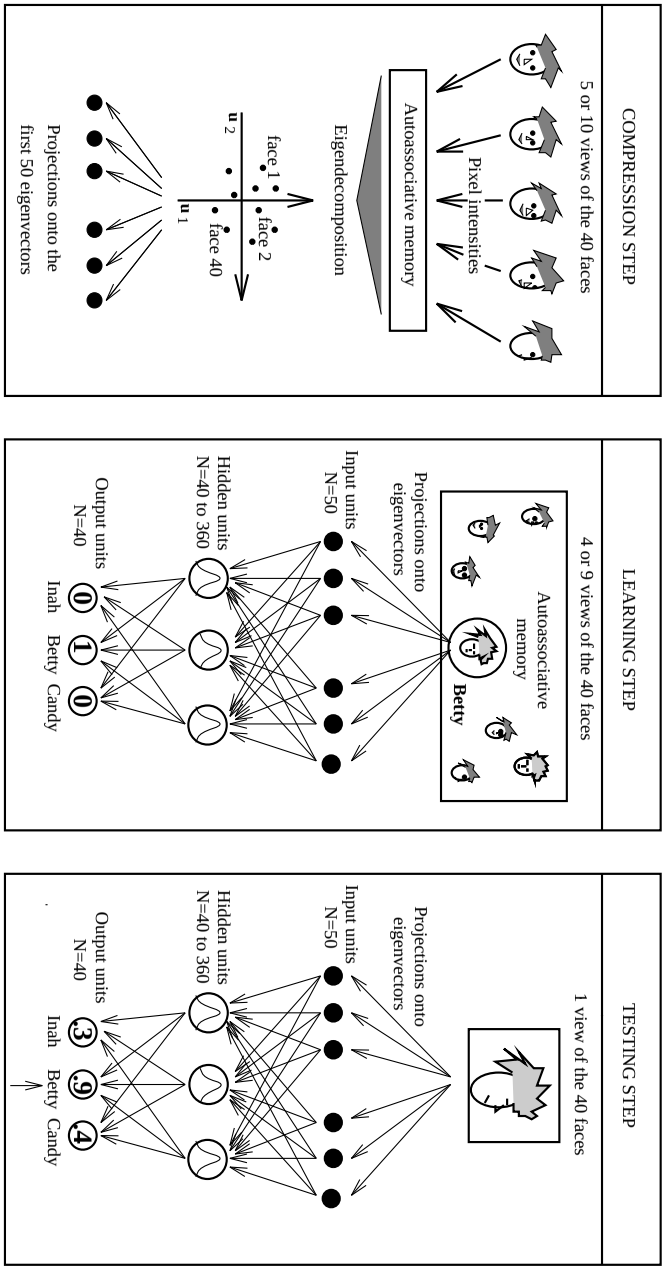

FIGURE 5. Illustration of the compression, learning and testing steps.

Figure 5, the face of "Betty" is presented as input, the network will be trained to associate the value of 1 in the output unit corresponding to "Betty" and 0 in all the other output units.

The variance for the Gaussian function of the hidden units (σ^2) was set to 1. Some preliminary simulations showed that increasing the variance does not improve the performance of the

FIGURE 6. Illustration of a face reconstructed with the first 50 eigenvectors of an autoassociative memory. The memory was trained with five views of 40 female faces. The views sampled the rotation of the head from full-face to profile with about 20 degree steps.

network and that, decreasing it reduces its ability to generalize. At the end of learning, the weights of the RBF network were fixed.

- *Testing step.* One view of each of the 40 faces—the view that was not used during learning—was used as input to the RBF network and the activation of the output units computed for each face. The level of activation of the output units was used as a classification criterion: The target view was identified as the face corresponding to the output unit with the highest activation (winner-take-all strategy.) For example, the target view presented as input in Figure 5 would be identified as a view of Betty's face because the level of activity of the output unit corresponding to "Betty" (.9) is larger that the level of activity of the output units corresponding to the faces of "Inah" (.3) or "Candy" (.4). If the target view is really a view of Betty's face, the network is said to have made a correct identification. If the target view was not a view of Betty's face, the network is said to have made an identification error.

To optimize the number of views available for testing the model, this procedure was repeated until each view of the 40 faces was used, in turn, as the testing view (i.e. using a jackknife technique.) Thus, a total of 50 simulations was carried out: Five types of views (0, 20, 45, 70, and 90 degrees from full-face) $\times$ five types of centers (all views; average view; frontal and profile views; frontal, 3/4, and profile views; and 3/4 view) $\times$ 2 learning conditions (5-view and 10-view conditions.) At the end of each simulation, the number of identification errors produced by the network was recorded.

5.2. *Results and discussion.* The ability of the model to generalize to new orientations was analyzed by examining both:

- The proportion of identification errors as a function of the type of centers of the RBF network and the number of views presented during learning;
- The repartition of identification errors as a function of the view presented at test.

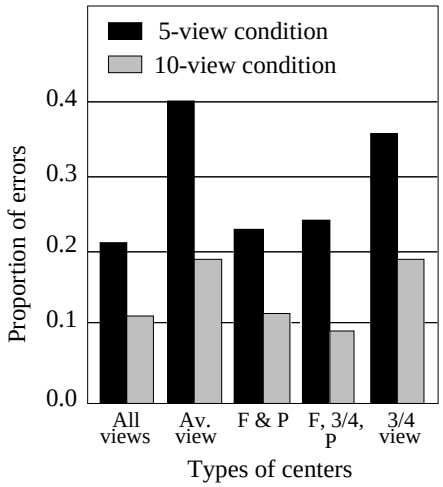

FIGURE 7. Proportion of errors as a function of the type of centers and the number of views presented during learning. **F** indicates a frontal view and **P** a profile view.

5.2.1. *Proportion of identification errors.* Figure 7 displays the proportion of errors as a function of the type of centers and the number of views presented during learning. Three points can be noted from this figure:

- The proportion of errors decreases when more views of the faces are presented during learning: The misidentification average is 30% in the 5-view condition, but only 15% in the 10-view condition.
- Performance depends on the type of center used: The largest number of misidentifications occured for the average view centers (40% error in the 5-view condition and 20% error in the 10-view condition) and the 3/4 view centers (36% error in the 5-view condition and 20% error in the 10-view condition.) No major differences were observed among the three other conditions (around 20% error in the 5-view condition and 10% in the 10-view condition.)
- There was no interaction between the number of learned views and the type of centers, which suggests that the optimality of the internal representation is not dependent on the number of views presented during learning.

The first conclusion that can be drawn from this series of simulations is that, having a 3D invariant internal representation is not necessary for a computational model to be able to *identify* faces from new orientations, nor is it necessary to store in memory all the familiar orientations of the faces. In fact, only two views (frontal and profile) are enough to reach 90% correct identification. The second conclusion is that in terms of distance from canonical or prototypical views, the internal representation suggested by single cell recording studies—two canonical views: full-face and profile—is more efficient than an internal representation involving only 3/4 views of faces. Moreover, the fact that, on the whole, no improvement was observed when a 3/4 view was added to the frontal and profile views suggests some optimality in the hypothesis derived from single cell recordings.

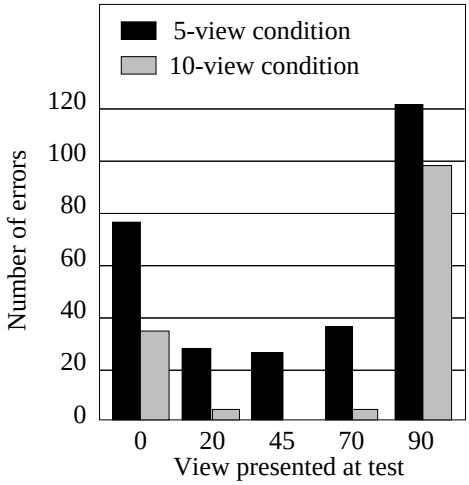

FIGURE 8. Repartition of identification errors as a function of the type of view presented at test. The horizontal axis represents the degrees of rotation from a frontal view.

5.2.2. *Error repartition.* Figure 8 displays the repartition of the identification errors averaged across the type of centers as a function of the type of views presented at test in the 5-view and the 10-view conditions. The main point to note from this figure is that the identification performance is better when an intermediate view (20, 45, 60, degrees from full face) is presented at test than when either a frontal or a profile view is presented. The worst performance is obtained in both the 10-view condition and the 5-view condition when a profile view is presented as a test view.

Although not surprising, this pattern of results is noteworthy because it replicates the 3/4 view advantage found with human subjects. Even more interesting is that this pattern of results is obtained with all the different types of centers (cf. Figure 9). This result indicates that the 3/4 view advantage cannot be interpreted as evidence for the memory storage of 3/4 views because this advantage is obtained even when the internal representation is only a frontal and a profile view.

6. Conclusion

The main results of the simulation are the following. First, the performance of the model improves with the number of views presented during learning for all types of representation. In the 5-view condition the model behaves somewhat like human subjects recognizing unfamiliar faces: It produces a lot of mistakes. In the 10-view condition, the performance of the model becomes closer to that of human subjects recognizing familiar faces: It becomes less sensitive to depth rotation. Second, in both learning conditions (5-view and 10-view conditions) the best performance was obtained when either 1) all the views, 2) a frontal, a 3/4, and a profile view, or 3) a frontal and a profile view were used as the centers of the RBF networks. The fact that the same performance was obtained in these three conditions indicates that the recognition/identification of a face from a new orientation can be achieved very efficiently by interpolating between two extreme orientations. Or, in other words, it is not necessary to store intermediate views in memory for a computational model to be able to recognize faces from new view orientations. Third, even when intermediate views are not used as the internal representation, they still yield the best performance. In other words, the fact

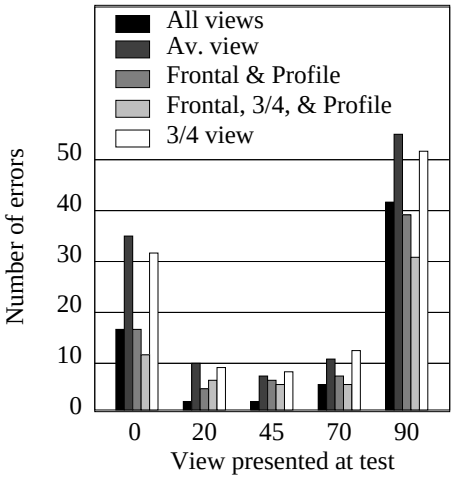

FIGURE 9. Repartition of identification errors as a function of the type of view presented at test and type of centers. The horizontal axis represents the degrees of rotation from a frontal view.

that a particular view of a face yields better performance does not imply that this view is stored in memory.

In conclusion, this simulation shows that the apparent contradiction between psychological and physiological data mentioned previously is not a real one. Recall that single cell studies found a larger number of cells specific to full-face and profile views than to 3/4 views, but that human data showed a strong advantage for 3/4 views. The simulation results reported here show this contradiction to be only apparent, because a 3/4 view advantage is obtained when only frontal and profile views are used as canonical or prototypical views. In fact, the pattern of results illustrated in Figure 7 suggests some optimality in the single cell hypothesis since performance was not improved when a 3/4 view was added.

This result suggests a different interpretation of the 3/4 view advantage than one previously proposed in the literature. Recall that the 3/4 view advantage has been interpreted as evidence that 3/4 views are stored in memory, and that recognition from other views is done by comparison to that view. An alternative explanation is that because the stimuli here sampled the rotation of the faces from frontal to profile view, recognizing a face from a 3/4 view is an interpolation task whereas recognizing a face from a profile or a frontal view is an extrapolation task. The main difference between these two types of tasks is the amount of available information: There is more information available to recognize a face from an intermediate view than from an extreme view. If this is the case, then, the relatively poor performance observed for frontal views should improve if multiple views of faces sampling the rotation of the head from left profile to right profile were used as stimuli. In this situation, since the same amount of information would be available for recognizing frontal and 3/4 views, the advantage of 3/4 views over frontal views should disappear. However, if 3/4 views are indeed canonical their advantage over frontal views would be maintained.

Appendix

.1. *Autoassociative memory.* An autoassociative memory is a network of I linear units or cells fully interconnected by way of modifiable connections or synapses. The connections between two cells i and j are bidirectional and symmetrical. The set of connections is represented by an $I \times I$ matrix $\mathbf{W}$ in which a given element represents the strength of the connection between two cells. To store a set of K faces in an autoassociative memory, the faces are first digitized and coded as pixel vectors, denoted $\mathbf{x}_k$, with each numerical element in $\mathbf{x}_k$ being the gray level of the corresponding pixel. For computational convenience, The vectors $\mathbf{x}_k$ are normalized so that $\mathbf{x}_k^T\mathbf{x}_k = 1$. The set of faces is represented by a $I \times K$ matrix denoted $\mathbf{X}$ in which the k-th column is equal to $\mathbf{x}_k$.

The K face images can be stored in the memory by setting the weights of the connections between cells using Hebbian learning.

$$\mathbf{W} = \mathbf{X}\mathbf{X}^T = \sum_{k=1}^{K} \mathbf{x}_k\mathbf{x}_k^T \tag{1}$$

where T denotes the transpose operation. Retrieval of a face is performed by presenting the face as input to the memory. Specifically, recall of the k-th face is achieved as

$$\widehat{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k \tag{2}$$

where $\widehat{\mathbf{x}}_k$ represents the answer of the memory. The quality of this answer can be estimated either by visually comparing the reconstructed face with the original face or, more formally, by computing the correlation or cosine between $\widehat{\mathbf{x}}_k$ and $\mathbf{x}_k$ (see e.g. Valentin, Abdi & O'Toole, 1994 for more details). The storage capacity of the memory can be improved by using a Widrow-Hoff error-correction learning rule:

$$\mathbf{W}_{[n+1]} = \mathbf{W}_{[n]} + \eta(\mathbf{X} - \mathbf{W}\mathbf{X})\mathbf{X}^T \tag{3}$$

where n represents the iteration step and η a small positive constant.

Since the weight matrix $\mathbf{W}$ is a cross-product matrix (and hence is positive semi-definite), it can be analyzed in terms of its eigen-decomposition as

$$\mathbf{W} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad \text{with} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{4}$$

where $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{W}$, and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. As a consequence, a learned face can be expressed as a linear combination of the eigenvectors of $\mathbf{W}$:

$$\widehat{\mathbf{x}}_k = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{x}_k \tag{5}$$

where the dot product $(\mathbf{u}_\ell^T\mathbf{x}_k)$ represents the projection (or weight) of face k on eigenvector $\mathbf{u}_\ell^T$. If complete Widrow-Hoff learning is used, Eq. 4 reduces to

$$\mathbf{W}_{[\infty]} = \mathbf{U}\mathbf{U}^T \tag{6}$$

and Eq. 5 can be rewritten as:

$$\widehat{\mathbf{x}}_k = \sum_{\ell=1}^{L} \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{x}_k \ . \tag{7}$$

.2. *Radial basis function networks.* RBF networks are 2-layer feed forward networks in which the input vectors $\mathbf{x}_k$ perform a nonlinear mapping of the input layer onto the output layer. This nonlinear mapping can be regarded as a recoding of the stimuli presented as input. Specifically, each hidden unit computes the radial basis function ϕ of the distance from the input vectors $\mathbf{x}_k$ to its center $\mathbf{c}$:

$$h_i = \phi(\|\mathbf{x}_k - \mathbf{c}_i\|) \tag{8}$$

where h_i is the output of the i-th hidden unit and $\|.\|$ denotes the Euclidean norm. A variety of radial basis functions can be chosen. In the simulations presented in this paper we used a Gaussian function:

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-x^2/2\sigma^2\} \tag{9}$$

where σ^2 represents the variance of the Gaussian function. Note that the constant term $(\sqrt{2\pi\sigma^2})$ is used to normalize the integral to 1 and can be omitted without any problem.

The outputs of the hidden units are then propagated to the output units of the network that integrates them in the following manner:

$$\mathbf{o}_j = \sum_i w_{i,j}\mathbf{h}_i = \sum_i w_{i,j}\phi\{\|\mathbf{x}_k - \mathbf{c}_i\|\} \tag{10}$$

where $w_{i,j}$ represents the strength of the connection between hidden unit i and output unit j. Using a matricial notation, the activity of the J output units for the K exemplars presented as input can be expressed as:

$$\mathbf{O} = \mathbf{HW} \tag{11}$$

where $\mathbf{O}$ is the $K \times J$ output matrix, $\mathbf{W}$ is the $I \times J$ weight matrix, and $\mathbf{H}$ is the $K \times I$ matrix of hidden units activities. Learning is achieved by modifying the connection weights so as to minimize the difference between the $K \times J$ expected output matrix, denoted $\mathbf{T}$, and the actual output $\mathbf{O}$. In general, the optimum values for the weights can be obtained by using a least-squares approximation. Formally:

$$\mathbf{W} \approx \mathbf{H}^{+}\mathbf{T} \tag{12}$$

where $\mathbf{H}^{+}$ represents the Moore-Penrose pseudo-inverse of $\mathbf{H}$ (see e.g. Abdi 1994 for a more detailed description).

References

[1] Abdi, H. (1988). "Generalized approaches for connectionist auto-associative memories: Interpretation, implication, and illustration for face processing," in *Artificial intelligence and cognitive sciences*, edited by D. J. (Manchester University Press, Manchester), pp. 151–164.

[2] Abdi, H. (1994). "A neural network primer," Journal of Biological Systems **2**, 247–281.

[3] 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.

[4] Baddeley, A. and Woodhead, M. (1981). "Techniques for improving eyewitness identification skills," Paper presented at the SSRC law and psychology conference Trinity College Oxford.

[5] Bruce, V. (1982). "Changing faces: Visual and non-visual coding processes in face recognition," British Journal of Psychology **73**, 105–116.

[6] Bruce, V. (1988). *Recognising Faces* (Lawrence Erlbaum, London).

[7] Bruce, V., Valentine, T., and Baddeley, A. (1987). "The basis of the 3/4 advantage in face recognition," Applied Cognitive Psychology **1**, 109–120.

[8] Bruyer, R. and Galvez, C. (1989). "The structural orientation of the mental representation of faces," Archives de Psychologie **57**, 259–269.

[9] Davies, G., Ellis, H., and Shepherd, J. (1978). "Face recognition accuracy as a function of mode of presentation," Journal of Applied Psychology **63**, 180–187.

[10] Desimone, R., Albright, T., Gross, C., and Bruce, C. (1984). "Stimulus selective properties of inferior temporal neurons in macaque," Journal of Neuroscience **8**, 2051–2062.

[11] Fagan, J. (1979). "The origins of facial pattern recognition," in *Psychological development from infancy: Image to intention*, edited by M. Bornstein and W. Keesen (Erlbaum, Hillsdale).

[12] 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.

[13] Krouse, F. (1981). "Effects of pose, pose change, and delay on face recognition performance," Journal of Applied Psychology **66**, 651–654.

[14] Logie, R., Baddeley, A., and Woodhead, M. (1987). "Face recognition, pose and ecological validity," Applied Cognitive Psychology **1**, 53–69.

[15] Macmillan, N. and Creelman, C. (1991). *Detection theory: A user's guide* (Cambridge University Press, Cambridge).

[16] O'Toole, A., Abdi, H., Deffenbacher, K., and Valentin, D. (1993). "A low dimensional represention of faces in the higher dimensions of the space," Journal of the Optical Society of America A **10**, 405–411.

[17] Palmer, S., Rosch, E., and Chase, P. (1981). "Canonical perspective and the perception of objects," in *Attention and performance IX*, edited by J. Long and A. Baddeley (Erlbaum, Hillsdale).

[18] Patterson, K. and Baddeley, A. (1977). "When face recognition fails," Journal of Experimental Psychology: Human Learning and Memory **3**, 406–417.

[19] Pentland, A., Moghaddam, B., and Starner, T. (1994). "View-based and modular eigenspaces for face recognition," IEEE Conference on Computer Vision and Pattern Recognition pp. 84–90.

[20] Perrett, D., Mistlin, A., and Harries, M. (1989). "Seeing faces: The representation of facial information in temporal cortex," in *Seeing contour and colour*, edited by J. Kulikowski, C. Dickinson, and I. Murray (Pergamon, Oxford), pp. 770–754.

[21] Perrett, D., Mistlin, 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 *Aspects of face processing*, edited by H. Ellis, M. Jeeves, F. Newcombe, and A. Young (Nijhoff, Dordrecht), pp. 187–198.

[22] Perrett, D., Mistlin, J., and Chitty, A. (1987). "visual neurons responsive to faces," Trends in Neurosciences **10**, 358–364.

[23] Perrett, D., Oram, M., Harries, M., Bevan, R., Hietanen, J., Benson, P., and Thomas, S. (1991). "Viewer-centered and object-centered coding of heads in the macaque temporal cortex," Experimental Brain Research **86**, 159–173.

[24] Perrett, D., Oram, M., Hietanen, J., and Benson, P. (1994). "Issues of representation in object vision," in *The neuropsychology of high level vision: Collected tutorial essay*, edited by M. Farah and G. Ratcliff (Erlbaum, Oxford), pp. 33–61.

[25] Perrett, D., Rolls, E., and Caan, W. (1982). "Visual neurons responsive to faces in the monkey temporal cortex," Experimental Brain Research **47**, 329–342.

[26] Perrett, D., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A., and Jeeves, M. (1985). "Visual cells in the temporal cortex sensitive to face view and gaze direction," Proceedings of the Royal Society of London B **223**, 293–317.

[27] Sirovich, L. and Kirby, M. (1987). "Low dimensional procedure for the characterization of human faces," Journal of the Optical Society of America A **4**, 519–524.

[28] Turk, M. and Pentland, A. (1991). "Eigenfaces for recognition," Journal of Cognitive Neurosciences **3**, 71–86.

[29] 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.

[30] Valentin, D., Abdi, H., and O'Toole, A. (1994). "Categorization and identification of human face images by neural networks: A review of the linear autoassociative and principal component approaches," Journal of Biological Systems **2**, 413–429.