PERGAMON

Vision Research 38 (1998) 2351–2363

Vision Research

Stimulus-specific effects in face recognition over changes in viewpoint

Alice J. O'Toole [a,*], Shimon Edelman [b], Heinrich H. Bülthoff [c]

[a] *School of Human Development GR4.1, University of Texas at Dallas, Richardson, TX 75083-0688, USA*
[b] *Department of Cognitive Science, University of Sussex, Brighton, UK*
[c] *Max Planck Institute for Biological Cybernetics, Tübingen, Germany*

Received 9 January 1997; received in revised form 8 January 1998

Abstract

Individual faces vary considerably in both the quality and quantity of the information they contain for recognition and for viewpoint generalization. In the present study, we assessed the typicality, recognizability, and viewpoint generalizability of individual faces using data from both human observers and from a computational model of face recognition across viewpoint change. The two-stage computational model incorporated a viewpoint alignment operation and a recognition-by-interpolation operation. An interesting aspect of this particular model is that the effects of typicality it predicts at the alignment and recognition stages dissociate, such that face typicality is beneficial for the success of the alignment process, but is adverse for the success of the recognition process. We applied a factor analysis to the covariance data for the human- and model-derived face measures across the different viewpoints and found two axes that appeared consistently across all viewpoints. Projection scores for individual faces on these axes (i.e. the extent to which a face's 'performance profile' matched the pattern of human- and model-derived scores on that axis), correlated across viewpoint changes to a much higher degree than did the raw recognizability scores of the faces. These results suggest that the stimulus information captured in the model measures may underlie distinct and dissociable aspects of the recognizability of individual faces across viewpoint change.

Keywords: Face recognition; Viewpoint change; Human; Model

1. Introduction

To recognize a face from a novel view, we must be able to encode something unique about the face that distinguishes it from all other faces in the world and, furthermore, we must be able to access this unique information from the novel view. Studying the representations and process that humans use to accomplish this task is difficult due to the complexity of the visual information observers experience in viewing faces from different viewpoints and due to the multitude of ways that such information can be encoded and represented.

We believe that to study human representations of faces, we must first study human faces, both as individuals and as exemplars of a category of objects that share a common physical structure. Our approach here is to combine human data on the recognizability and viewpoint generalizbality of individual faces with a computational model of the representations and processes required to perform this task. The rationale behind this approach is that individual faces vary both in the quality and quantity of the 'uniqueness' information they provide a human observer for the task of recognition. Some faces are highly unique and distinct, while others are rather less so. Not surprisingly, there is good evidence to indicate that faces judged by human observers to be unusual are more accurately recognized than faces judged to be typical (e.g. [1]). Despite the well-known findings relating the typicality and recognizability of faces, less is known about the ways in which faces can be typical or unusual. It seems reasonable, however, to suppose that when faces are considered distinctive, the information that makes them so can vary quite radically in its qualitative form, e.g. from very local features (moles, scares etc.) to much

* Corresponding author. Tel.: +1 972 8832491; e-mail: otoole@utdallas.edu.

0042-6989/98/$19.00
PII: S0042-6989(98)00042-X

more global deviations in head shape (the face of Meryl Streep). In the context of recognizing faces across a viewpoint change, there is no reason to believe that qualitatively different kinds of distinctiveness will transfer to novel views with the same efficiency, i.e. it may be that a distinctive head shape is visible and comparable from all views, but that an unusual facial feature is most salient from a single or small set of views. (While measuring the recognizability of individual faces is generally not the standard approach to studying face recognition, it is clearly not without precedent. This approach has been used successfully with frontal views of faces with the goal of relating face recognizability to other rated facial attributes such as typicality. In fact, the well-known link between face typicality and recognizability is an important part of the evidence for prototype theory for faces [1–4].) If indeed there are qualitatively different kinds of information in faces which transfer to novel views with different efficiency, then the complexity of these transfer patterns will be lost unless some consideration is given to the pattern of recognition and view generalizability for individual faces.

In this introduction, we present a brief overview of what we believe are the critical requirements of a generic model of human face recognition. We then extend these requirements to deal more specifically with the question of viewpoint generalization.

1.1. Computational models of human face recognition: the critical requirements

The past decade has seen a surge of interest in computational modeling of face recognition. In contrast to the complexity and diversity of models in the literature, we would argue that the basic common requirements of a computational model of human face recognition are somewhat simpler than the number of models proposed might suggest. Indeed, we believe that many current models meet our generic requirements albeit in different ways. Primarily, we believe models of human face processing must be sensitive both to the statistical structure of faces and to the statistical structure of our experience with faces.

The requirement that models be sensitive to the statistical structure of faces refers to the combined implications of two facts: (1) faces comprise a class of objects; and (2) face recognition requires the ability to distinguish amongst, and to remember a very large number of individual exemplars from within this object class. This makes the task of face recognition markedly different from most common definitions of object recognition, which typically involve categorizing exemplars as members of a particular object class. (We are not suggesting here that face recognition is 'special'. In fact, we believe that when the 'face recognition' definition applies to object recognition (e.g. we must recognize our car from among other similar models), the task constraints and model requirements would be comparable. In general, however, the difficulty of the problem in real life, as defined by the number of individual exemplars we must keep straight (i.e. how many individual chairs, cars, coats, suitcases do we have to keep track of?), is perhaps not completely comparable. We leave it as an exercise to the reader to try to imagine what object class, other than faces, requires us to keep track of individual exemplars, and then to consider, exactly how many exemplars are involved.)

Remembering a large number of individual faces is a difficult task because faces share a similar structure. Thus, the kinds of features that are likely to be helpful in most cases will be face-specific-for example, the distinctive feature defined by eyes that are 'two close together' may make a particular face very memorable. This feature is clearly meaningless for objects other than faces. Also, its usefulness in making a face memorable implies, of course, that one has an idea of how far apart the eyes usually are and hence, also has an idea of what constitutes 'too close together' or 'two far apart'. Our ability to use such a feature is an indication that we are indeed sensitive to the statistical structure of faces.

The second requirement, that the model be sensitive to the structure of individual observer experience with faces, refers to the effects that a lifetime of face learning may have on the representational system itself. The well-known difficulty we have recognizing faces of 'other' races by comparison to faces of our own race is an example of this. The phenomenon can be modeled by representing faces using a feature set derived from the statistical structure of a set of faces, varying the proportion of different races of faces [5].

1.2. The problem of viewpoint generalization

It is important to note that the large majority of computational models of face recognition operate on faces from a single, usually the frontal, view, though we mention a few exceptions below. Object recognition models, on the other hand, have taken as their primary concern the problem of 'recognition' across changes in viewpoint [6]. This almost certainly reflects the difference in the nature of the problem from recognition of individual exemplars for faces versus the categorization of exemplars into basic level categories for objects. The problem of combining the recognition of individual exemplars of faces with the complication of doing so across viewpoint change has been much less studied (some of the recent exceptions

Fig. 1. A full, three-quarter and profile view of a face.

are in psychophysical studies [7–9] and computational studies [10–14]).

We purpose that the combination of the within-class of the recognition task required for faces, and the requirement that face recognition operate at least somewhat successfully from novel views (even when only a single view has been seen), mandates a combination of approaches. The model we implement is based on a model proposed by Lando and Edelman [10] and operates by: (a) 'aligning' or transforming a view of a face into a learned view (cf. [15]. In the present work we consider only one of the possible alignment rubrics considered by Ullman. Especially, we make use of a match to a two-dimensional image model, rather than a three-dimensional model); and (b) by interpolating within a view-specific module to determine if the aligned face is 'known' or 'unknown' [16]. As such, the model combines image-based representations, which provide rich and complex perceptual information about the structure of faces, with general knowledge about image transformations that can be learned. This alignment process can enable access to some of the subtle structural information available in an image-based, view-dependent representation across a much larger range of views than is otherwise possible.

The rest of this paper is organized as follows. First, we present human empirical data from a recognition across viewpoint change experiment. These data replicate the standard pattern of view-point dependency found in previous studies. We then re-analyze the data to extract measures of the recognizability and viewpoint generalizability of the individual faces. Next, we collected human observer ratings of the typicality on the individual faces from different viewpoints. We then present an overview and application of the proposed model to the recognition and viewpoint generalization problem with the goal of extracting model-derived measures of the typicality, recognizability, and viewpoint transferability of the individual faces. Finally, we applied factor analysis to the covariance pattern of model and human measures on the faces. We interpret these data by comparison to simply controls that try to relate the recognizability of the faces across the different transfer conditions.

2. Experiment 1

The purpose of experiment 1 was to collect performance data on observers recognizing faces across a viewpoint change and on the recognizability of the individual faces over viewpoint change. However, for comparison with other experiments of this type, we first present a standard analysis of the results in terms of observer measures. We then analyze the data in terms of the recognizability of the individual faces across viewpoint change.

2.1. Method

2.1.1. Observers

Ninety volunteers roughly half male and half female, between the ages of 18 and approximately 45 years old, were observers in the experiment. All were recruited from the University of Texas at Dallas (UTD) staff and student population.

2.1.2. Stimuli

Seventy-two volunteers between the ages of 18 and approximately 40 years old from Tübingen, Germany volunteered to have their heads scanned by a Cyberware™ laser scanner. These produced a three-dimensional surface model of each head and a texture map, which is a standard RGB-image of the head that maps point-to-point onto the head model. Most of the hair was digitally removed from the laser scan model, leaving only a trace of the hair line in most scans. The face stimuli used here were made by taking right and left full, three-quarter, and profile views of the laser scans. Fig. 1 shows the full, three-quarter, and profile view of a face.

2.1.3. Apparatus

All experimental events were controlled by a Macintosh Power PC 6100 programmed using PsyScope [17].

2.1.4. Procedure

The experiment was a standard yes/no face recognition study varying the learned view (full, three-quarter, profile) and the test view (full, three-quarter, profile)

and measuring recognition performance as the sensitivity or *d'* for discriminating learned and novel faces in each of the nine transfer conditions. Observers were assigned randomly to one of the three learning conditions (i.e. 30 observers per learn condition), and were instructed that they would view a series of faces from a particular view and would be required to recognize the faces subsequently, possible from a different view. Thirty-six faces were presented, one at a time for 8 s each. Exposure time was set based on a pilot study that indicated that the task was difficult. Observers then viewed all 72 faces and responded 'old' or 'new' using labeled keys on the computer keyboard. The test face remained on the screen until the observer responded. Since we were interested in getting measures of the recognizability of individual faces in all nine conditions, counterbalancing was implemented to assure that *d'*s for individual faces in each condition could be based on an equal number of presentations of the face as old and new.

Additionally, after the completion of the recognition experiment, a subset of 30 of the observers was assigned randomly to view group (full, three-quarter or profile) (excluding the one that they learned in the recognition experiment). These observers (ratings can be made only with a single view condition, of which we had only three, hence the smaller number of observers) were asked to make a variety of facial ratings on all 72 faces, all presented from a single constant viewpoint. We consider only the typicality rating in the present study. Observers rated the typicality of the face on a scale of 1–3, with one being unusual and 3 being very typical.

2.2. Results

The hit and false alarm rates for each observer in each condition were assessed, and *d'*s were computed from these rates. These data were submitted to a two-factor (3×3) analysis of variance (ANOVA), with the learned view (full face, three-quarter, and profile) as a between subjects factor and the test view (full face, three-quarter, and profile) as a within-subjects factor. The main effect of the learned view was found, $F(2, 86) = 11.48$, $\mathrm{MS} = 4.78$, $P < 0.0001$. A smaller, through significant effect of the test view was also found, $F(2, 172) = 3.41$, $\mathrm{MS} = 1.32$, $P = 0.3$. Finally, both of these main effects were qualified by the highly significant interaction between the learned and the test view, $F(4, 172) = 23.34$, $\mathrm{MS} = 9.31$, $P < 0.01$. The pattern of interaction means is displayed in Fig. 2.

For reasons that will become clear with reference to our analysis of the recognizability of individual faces, we also computed an ANOVA on the criterion used by the observers as a function of the learned and the test condition. This yielded a highly significant interaction between the learned and the test view, $F(4, 172) = 17.9$, $\mathrm{MS} = 3.26$, $P < .0001$. The pattern of interaction means were relatively clear-negative or loose criteria were found in the no-transfer conditions, positive or strict criteria were found in all transfer conditions. Thus, observers responded 'old' much less frequently in the transfer conditions than in the no-transfer conditions [18].

2.3. View condition measures on faces

2.3.1. Procedure

Just as for the individual observers, the discriminability of the individual faces can also be assessed using standard signal detection theory measures. (An extensive analysis of the theoretical interpretations and procedures used in this kind of analysis applied to faces in a face recognition experiment can be found in [19].) One proceeds by collapsing the data across observers and computing a *d'* for each face. This was done for each face in all nine learn–test conditions of the experiment. Thus, we assessed hit and false alarm rates for each face in each condition by compiling data across different observers. The method of computing the hit rates for individual faces was straight forward. An example serves to illustrate. When an observer responded 'old' to a given face that was learned as a full face and was tested as three-quarter face, a hit was recorded for that face in the full to three-quarter transfer condition. Hit rates for all faces in all transfer conditions were computed likewise.

The computation of a false alarm rate for individual faces was more complicated as a 'new' face does not have a learn condition. There were two possibilities for computing the false alarm rate. The most obvious was to only compute a 'test' false alarm rate, yielding three (full, three-quarter, and profile), rather than nine, false alarm rates for each individual face. Using this proce-

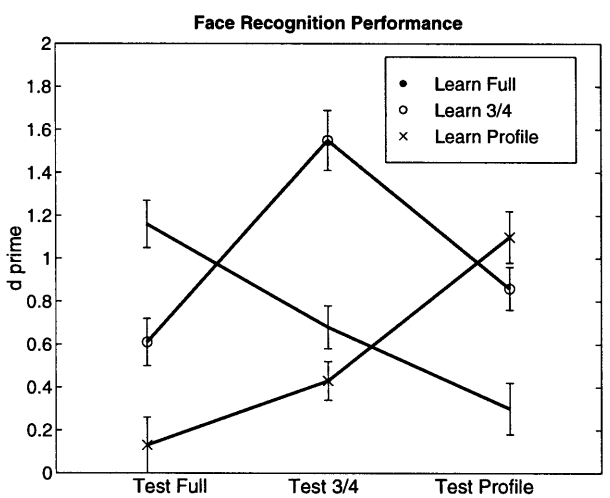

Fig. 2. Face recognition accuracy data as a function of learned and test pose in experiment 1.

dure the false alarm rate for a face in the three transfer conditions, e.g. full to full, three-quarter to full, and profile to full, is defined as the number of times observers responded 'old' to a novel face, e.g. tested as full. A second possibility was to impose the learn condition of the observer on the 'novel' face. Thus, observers who learned all faces as full, would provide the false alarm rates for faces in the full to full, full to three-quarter, and full to provide transfer conditions.

We chose the latter method due to the observer-based criterion variations that operated in different transfer conditions. The counterbalancing scheme we employed assured us that each of the 72 faces appeared equally often as a lean and test face across the 90 observers. Within each condition, across the observers, each face appeared a total of 10 times. In summary, the data resulting from this analysis consisted of nine *d'*s for each of the 72 faces, one for each of the transfer conditions.

2.4. Discussion

The human empirical data give an interpretable and relatively standard picture of face recognition across viewpoint change and are consistent with the previous work indicating viewpoint dependency. Additionally, the results replicate the well-known recognition advantage for 'three-quarter' view faces. This three-quarter advantage cannot be accounted for solely on the basis of the fact that it is the center view of the three views tested. This is evident from a comparison of the no transfer conditions for the full, three-quarter and profile faces, which shows that the performance for no transfer three-quarter conditions was significantly better than for the no transfer profile and full conditions (cf. error bars in Fig. 2).

In summary, the experimental data replicated the 'three-quarter' advantage and viewpoint dependency found in previous studies and provide us with *d'* measures for the individual faces in each of the nine view transfer conditions. We also have a measure of the typicality of each face from each of the three viewpoints.

3. The computational model

The purpose of the computational model was to provide measures of the recognizability and viewpoint generalizability of individual faces. By contrast to the human generated measures, the model performance measures are based purely on the physical properties of the stimuli under the coding and processing assumptions we make. In other words, these performance measures for each face are determined by the quality and quantity of the information the model can exploit for recognizing and generalizing the face across viewpoint. We implemented a slightly altered version of the model described in [10]. Because that paper provides a detailed description of the model, and because its theoretical underpinnings are discussed elsewhere [20], we present only the essentials of the model interpretation and implementation in the present paper and refer interested readers to these other papers. In the present paper, we focus on the qualitative predictions relating typicality, recognition and viewpoint generalizability.

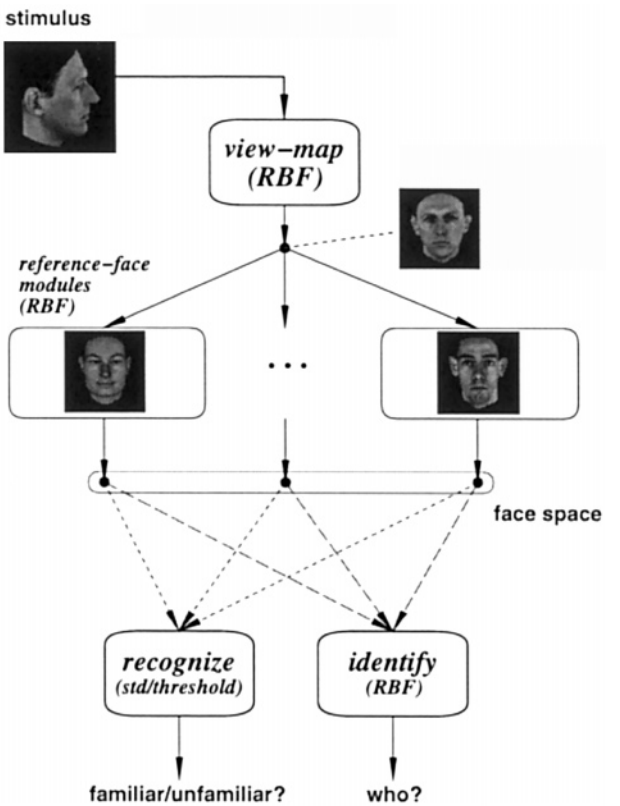

Fig. 3. Schematic overview of the model showing the output of the view-mapper process, and the interpolation in the canonical view modules.

The essential components of the computational model consist of a view alignment procedure and a recognition by interpolation procedure (Fig. 3). (Like all models that make use of alignment, we assume that the viewpoint of the face is detectable (cf. Lando and Edelman for a discussion of this issue). We assume also that view alignments are done to at least a subset of canonical views, though we have no evidence for this other than suggestive physiological data indicating face-specific cells with viewpoint preferences (reviewed in [21]).) In the next two sections, we will walk through the implemention of the model using the figure as a guide.

3.1. View alignment procedure

The problem of recognizing a novel view of a face is handled by mapping the novel view onto other views, one of which may have been learned. This is illustrated schematically at the top of the figure, with the profile view of a face tranformed to its frontal view. This is a class-based transformation that the model learns from example face view transformations. (The transformation was approximated by the average transformation in [10] and hence the present implementation represents a substantial improvement in implementation.) By 'class-based', we mean simply that the mechanism learns the transformation using only faces, and is therefore, face-specific. This is one way in which the model is sensitive to the statistical structures of faces, or more precisely, to the relationship between the appearance of a face from two views. The psychological claim here is that when we see a completely novel face from a single view, we can infer what it might look like from another view. (The inference here is meant as in Helmholtz's "unconscious inference"; people are notoriously bad at consciously imaging what objects look like from unfamiliar viewpoints [22].) Although we may not be completely accurate in all or even many cases, it is certainly clear that we are more than competent at matching two views of a face and distinguishing this match from an alternative face [8].

3.2. Recognition by interpolation procedure

The recognition procedure is carried out independently within view modules. In other words, different modules deal with different views. When no view alignment is required, (the face is learned and tested from a single viewpoint), the input to the recognition test consists of a face image from the appropriate view. When view alignment is required, (the test face is presented in a novel view). The input to the recognition stage consists of the output of the appropriate view mapper. This output is simply the view mapper's estimate of the appearance of the face image from the appropriate view.

Within each view module, the recognition process makes use of a two-layer radial basis function (RBF [16]) network as the main computational mechanism. At the first layer, the original or view mapped input face is compared to a set of reference faces, by computing the outputs of a set of RBF modules, each of which has been trained previously on one of the reference faces. This can be seen in Fig. 3 between the view mapper and the level marked 'face space'. The similarity profile of the face with respect to the reference faces is the representation of the face.

The reference faces are randomly selected faces that serve as a coding base for any other face that may input to the system [23]. The use of reference faces as a coding basis is simple a way to define a face space via examples of faces, and hence to define new faces with respect to a multidimensional measure of their similarity to the example faces. We refer to the activation of the output units at this layer for a given face as the projection of the face into the face space. The face space representation of the stimulus is then mapped to a second layer of RBF units that represent the identity of the face. This is illustrated at the bottom of Fig. 3. In the present paper, we consider only the 'recognition' output of the model.

Translating this relatively technical description of a face space into more intuitive terms, we wish to be explicit about the nature of this representation of faces and how it differs from other representations. First, the representation of a face in terms of a set of reference faces explicitly contrasts a representation by similarity to a representation of similarity [20]. A representation by similarity makes use of the similarity between a stored face and a test face to determine if the test face is known. In other words, analog representations of the face images comprise the face space. This works well only when learned and test face encodings are similar. A representation of similarity makes use of the similarity structure of a face space, as the representation of a face. This works well both when learned and test face encodings are similar, but also, when only the similarity relationships among the learned and reference faces are preserved (after some transformation, for example) This is a very important difference. It means that the class-based transformation has to preserve only enough information to retain the similarity structure of the face with respect to the references faces. This is considerably easier than using the similarity between the face and a single template as the only available basis for the recognition decision.

3.3. Methods

The model was implemented as described in [10] with only a small improvement in the implementation of the alignment process, which we describe below.

3.3.1. Stimuli

The same 72 faces used with the human observers served as the model stimuli. The learning procedure we implemented required that these 72 faces be divided into four sets of 18 faces each. Two of these sets formed the basis of a 'longer-term' perceptual and memory component (i.e. the face used to train the view mapper and the faces used as reference faces for the first network layer). The second two sets were used as the experimental learned and test faces. Due to the small number of faces that remained per group, in all cases, the simulation data consisted of a full counterbalanced

set of 24 simulations, with different sets of faces serving equally often in each part of the process, (i.e. view map training set, reference face set, learn/target face set, and test/distractor face set).

The face images, which were originally 384×384 colour pixels, were reduced to 96×96 and converted to a grey level. Prior to the simulations, the mean pixel value was subtracted from each image and the images were histogram-equalized. (We carried out a limited number of simulations using faces that were less reduced in size but found little improvement in the model performance. Additionally, we note that the histogram equalization was implemented for technical reasons having to do with the potential sensitivity of the RBFs to the pixel distributions. In any case, the effects of this choice are on the conservative side making it harder for the model to fit the human data.) It is worth noting that although we believe strongly in the importance of image-based codings for this model, raw pixel maps are a simplification. Other 'elaborated' image-based codes, including pre-processing similar to that used for morphing, might be preferable.

3.3.2. View Normalization

Simple linear associative view mappers were created for all combinations of the views transfer conditions (full, three-quarter and profile). Each was trained with 18 face pairs to produce an output view of a face when the input view was presented—thus, for example, the full to three-quarter mapper learned to produce the three-quarter view of the face when the full face view was input. The faces use to train the view mapper were not used elsewhere due to the fact that the linear associator performs the view map perfectly for the faces on which it is trained. The role of the view mapper was to approximate the transformations for unknown faces based on its general knowledge about faces.

3.3.3. RBF mapping to the face space

For brevity and concreteness, we describe the model particulars for a single view pair condition (learn full-face and test three-quarter face). All view conditions were learned in the same fashion. Eighteen additional full faces, not used to train the view mapper, were used to train the RBF modeules corresponding to the reference faces (that is, the modules that mapped the input into the face space). Codings at this RBF layer, therefore, captured the similarity structure among the reference faces. The RBF activation patterns were mapped to an output layer containing 1 unit per face. Target vectors for the reference faces were strings of 18 zeros, with a single 1, uniquely placed for each different face.

3.3.4. RBF mapping to the recognition output space

A third set of faces, in combination with the face set used to train the previous network, was used to train a final network mapping from the face space to an identity space with 36 output units. (Half of these 36 units were devoted to the reference faces and half to the learned faces, so that the model would have an identity level coding at this second layer output for both sets of faces.) This third set of faces was considered the 'learned' face set.

3.3.5. Test faces

The fourth set of faces was reserved as novel or distractors face set.

3.3.6. Procedure

The four sets of faces were used in all possible combinations ($n = 24$), with each face set appearing 6 times as the learned set and 6 times as the novel set. All 24 of these simulations were carried out for each of 6 transfer conditions between face views: (1) learn full-test three quarter; (2) learn full test profile, etc.

3.4. Model Testing

We implemented this model to extract data on the 'performance' of the individual faces at both the view alignment and recognition stages.

3.4.1. View alignment performance

We measured the success of the alignment process for each face in terms of the similarity between the estimated and actual face in the reconstructed view. This measure was simply the cosine between the face vectors for the original and estimated views. We did this for each face in all possible transfer conditions

3.4.2. Interpolation or recognition performance

To recognize faces the model must be able to distinguish between learned and novel faces using some aspect of the output activations. The standard psychological theory of the task suggests that we are generally able to recognize faces because known faces evoke stronger feelings of familiarity than unknown faces. A yes/no recognition decision is made by setting a familiarity criterion such that faces that evoke familiarities higher than this criterion are categorized as 'old', whereas faces that evoke familiarities less than the criterion are rejected as 'new'. As for the observers, correct recognitions or 'hits', are old faces, correctly called 'old'. False alarms are new faces, incorrectly called 'old'. From these hit and false alarm rates we computed standard discrimination indices. The measure of familiarity we used was the S.D. of the output unit activations for old versus new faces. Old faces produced more 'differentiated' output patterns, with relatively stronger excitations for the target pattern, and greater inhibition for the nontarget patterns. (It is worth mentioning briefly that because the view–map process is

not needed for the nontransfer conditions, like many other computational models of face recognition that do not explicitly add noise to the test face (e.g. principal component based models), the present model performed perfectly on faces that were learned and tested from the same view.)

Using this familiarity measure the recognition performance of the model appears in Table 1. Because the model makes available the entire old and new distributions, recognition was measured as area under the ROC curve. As can be seen, the model is capable of good though not perfect recognition performance and recognition performance declined for the larger viewpoint change.

4. Combining model and human data

The human empirical data analysed at the level of observer performance gave a relatively standard and interpretable pattern of viewpoint dependency in face recognition. But how well do the observer-based measures characterize recognition and viewpoint generalization performance at the level of individual faces? More precisely, to what extent do the representations we make of faces from different views relate to one another? A very simply way to ask this question quantitatively is to correlate the recognizability of the faces across the different viewpoint conditions. We did this using the 'recognizability' scores (d's) that we computed for each face in each of the nine transfer conditions. For simplicity, we begin with correlations among the no-transfer conditions (e.g. full to full). These appear in Table 2. Although statistically significant, the intercorrelations of face recognizability among the no-transfer conditions were weak, explaining only 17% of the variance in the very best case. The small amount of variability accounted for indicates that the raw recognizability scores of the faces do not provide a strong linkage between human performance on the individual faces in the different viewpoint conditions.

We next calculated these correlations for the transfer conditions and found that the correlations were even weaker than for the non-transfer conditions (Table 3), explaining only 14% of variance in the best care. We present only a subset of the 36 possible correlations—those with a common learned view (e.g. full to full, full to three-quarter, and full to profile). We expect these to be the most related among the transfer conditions. (Visual inspection of the associated scatterplots of recognizability between faces in pairs of conditions indicated that these low correlations were not due to some other nonlinear pattern relating the recognizability of faces, but reflected the general lack of a systematic relationship between the recognizability of faces under different conditions.)

These data indicate that the recognizability scores of individual faces in the different view transfer conditions are only weakly related. A simple interpretation of this result is that faces that are well-recognized in a particular viewpoint change condition are not necessarily well recognized in other viewpoint conditions. Although these results may seem surprising at first, with more consideration of the complexity of the information in faces and the complexity of the tasks involved in accessing this information from different viewpoints, the re-

Table 1
Summary of the model's performance in the recognition task

Learn	Test		
	Full	3/4	Profile
Full	—	0.85	0.77
3/4	0.84	—	0.87
Profile	0.78	0.85	—

The entries are areas under the ROC curve.

Table 2
Correlations between the recognizability of faces in the no-transfer conditions

View combinations	Correlation coefficients
Learned full and tested full correlated with learned 3/4 and tested 3/4	0.35 ($P<0.01$) $r^2=0.12$
Learned full and tested full correlated with learned profile and tested profile	0.41 ($P<0.01$) $r^2=0.17$
Learned profile and tested profile correlated with learned 3/4 and tested 3/4	0.34 ($P<0.01$) $r^2=0.115$

Table 3
Correlations between the recognizability of faces in transfer conditions when full-face was learned (top), 3/4-face was learned (centre), and profile-face was learned (bottom)

View combinations	Correlation coefficients
Transfer-condition: full-face learned	
Learned full and tested full correlated with learned full and tested 3/4	0.25 ($P<0.05$) $r^2=0.06$
Learned full and tested full correlated with learned full and tested profile	0.18 (ns) $r^2=0.03$
Transfer-condition: 3/4-face learned	
Learned 3/4 and tested 3/4 correlated with learned 3/4 and tested full	0.36 ($P<0.01$) $r^2=0.13$
Learned 3/4 and tested 3/4 correlated with learned 3/4 and tested profile	0.26 ($P<0.05$) $r^2=0.07$
Transfer-condition: profile-face learned	
Learned profile and tested profile correlated with learned profile and tested full	0.18 (ns) $r^2=0.03$
Learned profile and tested profile correlated with learned profile and tested 3/4	0.34 ($P<0.01$) $r^2=0.14$

sults are less surprising. In fact, a strong relationship between face recognizability across the view change conditions is surprising only if one assumes a homogeneity of the nature of the information in faces, such that this information is transfered in an equally efficient fashion across viewpoint change. In this case, the transfer problem would be constrained only by the degree of view change. The correlation data we present here indicated that this simple unidimensional model is not adequate to account for human performance at the level of individual faces.

Barring this unidimensionality, we applied a multidimensional factor analysis to the model- and human-generated face data with the hope of being able to establish a better linkage between faces across the different view conditions. We thought this is a reasonable expectation due to the fact that the model supplements the human data with face measures derived directly from a physical representation of the stimuli. It also provides us with a measure of face performance on the view alignment process that is independent of the ultimate success of the recognition procedure. As we will note shortly, there is reason to expect that the view alignment and interpolation processes will be differentially impacted by the typicality of faces. With factor analysis it is possible to dissociate faces on the basis of the relationships they induce in the potentially multidimensional pattern of model- and human-derived measures.

Factor analysis is a technique used to describe a set of correlated variables using a smaller number of uncorrelated or orthogonal variables (i.e. factors, axes). The factors, and hence, patterns of performance they dissociate, are ordered according to the proportion of variance they explain in the covariation of the face measures. We used this analysis here as a tool for describing the pattern of interactions among our measures, rather than as a tool for reducing the dimensionality of the data. To offset the inherent limitations of factor analysis as a non-inferential statistic, we conducted extensive Monte Carlo simulations on the loading patterns to test their stability. We limit our interpretations to those loadings that are stable with respect to the statistics collected in these simulations.

4.1. Model predictions concerning typicality

Because the view mapper is trained with a set of example faces, the performance of the view mapping process for a particular face provides a measure of the typicality of the face with respect to the view transform applied. The quality of view estimates should be better for faces that are 'typical' rather than unusual with respect to this transform, i.e. the view mapping procedure succeeds in so far as the face is close to the average (typical) and can be approximated in the new view with general information extracted from a set of learned faces. Although this is an unusual way to measure typicality we think that it is reasonable. Typicality is simply the extent to which simple operations on individual faces can be approximated by general knowledge about faces.

The beneficial effects of typicality on the view map procedure set up a paradoxical situation with the recognition problem. Typical faces, which are likely to be the most accurately view mapped, are not necessarily expected to be the easiest to recognize. Typical faces, once view mapped, are likely to be more similar to, and hence confusable with, other faces than are unusual faces. The face confusability factor is directly tapped in the second part of the model, the interpolation process, which is sensitive to the similarity relations among faces. Thus, relatively unsuccessful view maps (e.g. for the unusual face), do not necessarily lead to poor recognition, because the face, even badly approximated, may have few similar competitors vis a vis the similarity structure of the faces. The important point is that recognition can involve trade-offs between the success/failure of these two factors.

4.2. Procedure

The two model measures were as follows. The quality of the view mapper output was measured in term of the similarity between the estimated and actual face in the reconstructed view. We did this for each face in all possible transfer conditions and refer to the measure as the model's typicality estimate of the face. The second model measure was the recognition measure, which we defined as the difference between the S.D. in the activation of the RBF output units when the face was old versus new. This is a measure of the confusability of a learned face with the distractor faces. The two human empirical measures for each face were defined previously and are the face's typicality rating from each viewpoint and the face's recognizability.

For simplicity, and also due to the learning effect found in the human empirical results, we grouped the data according to the three learning views. This was also reasonable due to the fact that the face presentations on which human observers base view generalization judgements must be 'created' from the learning view. We then concentrated only on the conditions in which there was a view change, because the model performance for individual faces was perfect for the no-transfer conditions. Working at the level of the learning view, we computed one measure for each of the four variables (human and model recognition, human typicality rating, and model view map quality), by averaging the measures in the two transfer conditions within that viewpoint. For example, within the full viewpoint, we averaged the full to three-quarter and full

Table 4
Factor 1 for each of the three learned view analyses shows that the model and the human recognition measures agree and oppose the human typicality ratings and the quality of the view map

Learn	Factor 1		
	Full	3/4	Profile
Human recognition	**0.42**	**0.62**	**0.61**
Model recognition	**0.72**	**0.32**	**0.61**
Human typicality	**−0.70**	**−0.74**	**−0.30**
Model view map	**−0.45**	**−0.50**	**−0.58**
Prop. variance	0.35	0.32	0.30

to profile conditions for all measures except the human typicality rating, which was of course already specific to a single view. We then applied factor analysis separately to each of the three learn view condition matrices.

To test the stability of the factor loadings, we compared them with statistics (mean loadings and their S.E.) collected from 100 analyses that we conducted on permutated versions of the data. The 100 Monte Carlo simulations were carried out for each view condition. Each simulation was conducted as follows. First, we generated a random permutation of the original data matrix within the variable columns to ensure that the data distributions in the randomized simulations and in the analyses presented were identical. Second, we applied factor analysis to each permutated matrix. Finally, we computed the means and S.E. of the simulation loadings. In the factor tables that follow, we present, in bold-face, loadings that differed statistically from those computed in the randomized simulations (those greater than the corresponding simulated mean plus one S.E.).

We report first the general results of the factor analysis, in terms of the patterns of performance that were consistent across the viewpoint analyses. We then report the results relating the individual faces across viewpoint changes with respect to the axes.

4.3. General factor analysis results

To begin, in all three viewpoint analysis, the first axis showed a combined loading of the model and human recognition measures, opposing a combined loading of the human typicality rating and the model view map quality (Table 4). This factor can be described as follows. First, the combined loading of the model and human recognizability measures in the same direction indicates that to a first approximation the model and human-derived measures of face recognizability were related. Additionally, the model view map measure and the human typicality measures were also related. The opposition of the two sets of measures is consistent with the interpretation that faces well-recognized by the model and by human observers were judged atypical by observers and were not well estimated in the view mapping process (and vice versa). (The signs of the loadings are arbitrary, only the relationships among the signs are meaningful.) The human part of this axis simply picks the well-known inverse relationship between perceived typicality and face recognizability (e.g. [1]). The model measures of recognizability and view map quality form a computational, stimulus-derived, complement to this well known finding for human observers.

A second factor that appears consistently in all the three factor analyses is displayed in Table 5. This was the second factor in the full-face view analysis and the third in the three-quarter and profile analyses. (For brevity in the text, we will henceforth refer to this as the 'second consistent factor'.) The common feature of the axis is combined loading of the human recognition and the model view map quality in the same direction, rather than in opposition as was found for the first axis. The same-direction loading of the human recognition measure and the model view map indicated that the view mapper approximated these faces reasonably well, but that human accuracy was, nonetheless, good (and vice versa). This runs counter to the pattern noted on the first axis for which view map quality opposed human recognition of the faces.

Finally, the remaining two axes could be categorized only somewhat consistently across the three conditions and will not be discussed further.

In summary, two axes were comparable across the three separate viewpoint condition analyses. Combined they accounted for 57, 55, and 54% of the variance in the full, three-quarter, and profile analyses, respectively. The first of these axes indicated that model- and human-derived measures of typicality and recognizability inter-related in a straightforward way. In general, faces well recognized by the human observers in the different viewpoints were well recognized by the model, and faces rated typical by the model were accurately view mapped by the model. The second consistent axis indi-

Table 5
This factor shows a combined loading of the human recognition measure and the model view map measure in the same direction

Learn	Factor 2		
	Full	3/4	Profile
Human recognition	**0.78**	**0.67**	**0.67**
Model recognition	−0.01	−0.33	**−0.23**
Human typicality	0.07	0.01	0.28
Model view map	**0.59**	**0.54**	**0.61**
Prop. variance	0.22	0.23	0.24

Table 6
Correlations between the projection scores for individual faces on the factors extracted from the combined human and model data in the F, T and P learn conditions

Learn	Correlations	
	Factor 1	Factor 2
Full-3/4	**0.88**	0.11
3/4-Profile	**0.46**	**0.74**
Full-profile	**0.30**	0.14

Correlations significant at the 0.01 level appear in bold.

cated a less intuitive relationship between the model view map success and human recognition performance. We explore this in more detail in the context of relating these axes across three viewpoint analyses.

4.4. Relating individual faces across the viewpoints

The factor analysis combined model and human data in the three viewpoints and revealed two reasonable consistent axes across all three viewpoints. Our next question concerns the extent to which the pattern of face measures captured in these orthogonal axes can enable us to relate the performance of individual faces across these viewpoints. As noted, the pure recognition scores were not sufficient to do this. With the factor analysis we were able to supplement the psychological measures with information about the physical properties of the faces.

To formally assess the relationship of the faces across the viewpoint factor analysis, we proceeded as follows. First, we measured the projection scores of the faces onto each of the two consistent axes in each of the three viewpoint analyses; this yielded 6 projection scores per face. Intuitively, the projection score of a face onto one of the axes is simply a measure of the similarity of the face performance profile (defined by its two model and two human scores) to the performance profile captured by the axis. By spatial analogy, one can imagine that the two axes in each view analysis comprise a subspace, and that faces have coordinates in the subspace. The projection scores are simply the coordinates of the face with respect to the axes defining the subspace.

Using the face projection scores, we assessed the correlations between all possible view pairs on each axis. Specifically, for the first axis, we correlated the projection scores of the faces onto this first axis between: (1) the full and the three-quarter views; (2) the full and profile views; and (3) the three-quarter and profile views. These data appear in column 1 of Table 6. In contrast to the generally weak correlations relating the raw recognizability scores, the correlation of the face projection scores on the first factor provides a much stronger link between the face performance in the different viewpoints. The strongest relationship ($r = 0.88$) was found between the full and three-quarter learn conditions; a lesser correlation ($r = 0.46$) was found between the three-quarter and profile conditions; and the weakest correlation ($r = 0.30$) was found between the full and profile conditions. This would suggest that the information captured by this factor is highly reliable for transferring between the full and three quarter views; moderately reliable for transferring between the three-quarter and profile views; and only weakly reliable for transferring between the full and profile views.

We repeated the previous analysis for the second consistent axis. Theses data appear in column 2 of Table 6. A rather different pattern was seen here. A strong, reliable correlation ($r = 0.74$) occurred only between the three-quarter and profile condition. These data indicate that the information captured by this axis is most consistent in relating faces between the three-quarter and profile views.

5. Conclusions and general discussion

Combined, the consistency of two orthogonal factors in all three analyses indicates that the model and human data at the level of individual faces cannot be accounted for by a unidimensional model of the transfer process from the different views. This is consistent with: (a) the generally weak relationship we found between the human derived measure of face recognizability across the different transfer conditions; and (b) the simple interpretation that faces well-recognized in a particular viewpoint change condition are not necessarily well recognized in other viewpoint change conditions. The supplementary model-derived face information included in the factor analysis accomplished two things. First, it parcelled the human recognizability score into two relatively consistent components in all view conditions. Specifically, the human recognizability score loaded significantly on both of the axes that appeared consistently across all three viewpoint analyses. Second, we were able to use these components to dissociate faces based on the pattern of the model- and human-derived measures they showed. By coding the faces in terms of their adherence to these patterns we were then able to demonstrate much stronger results relating the faces across the viewpoints. One of the two axes related faces best between and the full and three quarter views, and the other axis related faces best between the three-quarter and profile views. This is consistent with the conclusion that although the recognizability of faces may be only weakly related across the viewpoints, the physical information in faces that differentially supports different viewpoint transfers may be more consistent. This information

may provide a foothold for understanding the distinct patterns of the relationships between the information in faces and how face recognizability is related across the viewpoint conditions.

What this analysis cannot provide is a precise picture of exactly what information underlies the transfer patterns captured in the two factors we found. As for all multidimensional factor analyses, it is possible to attempt a subjective interpretation of this information by locating individual stimuli with a very large positive and negative projection on each axis and comparing them. We have, of course, done this and speculatively suggest that the first axis captures global aspects of distinctiveness and the second factor captures locally distinctive features. This suggestion is based on our impression of what the faces at each end of the axes look like and on what we think the different patterns of model and human measures mean. More specifically, for this latter, although the model and human measures relate in an intuitive way on the first axis, the second axis actually provides a better logical foundation for interpretation. This is due to the counter-intuitive direction of the human recognizability and model typicality measures. The same-direction loading of these two measures indicates that the view mapper approximated these faces reasonably well, but the human accuracy was, nonetheless, good (and vice versa). Intuitively, these might be faces that were recognizable based on the presence of a relatively small distinctive feature. For the particular model we implemented and tested, mistakes on transforming these smaller distinctive features are likely to have a negligible impact of the quality of the view map measure. For human observers, however, small local distinctive features may be a valuable cue for recognizing and transferring faces across viewpoint change.

Interpretations of this sort, however, are only speculative. More work to analyze, very specifically, the physical structure of the faces as a function of their projection score values on the two axes would be needed to come to firmer conclusions on this matter.

We have argued that to understand the representations and processes that humans use to recognize faces across viewpoint, one must first study human faces, both as individuals and as exemplars of a category of objects that share a common physical structure. Our approach here has been to combine human data on the recognizability and viewpoint generalizbality of individual faces with a computational model of the representations and processes required to perform this task. The computational model we implemented represents a compromise between the need to retain the complexity of the perceptual information in faces and the need to have a representation flexible enough to generalize at least somewhat to new views. Although the present data suggest a theoretical reorientation toward the importance of studying the individual stimulus, the problems associated with doing so are not trivial. We are keenly aware that the methods needed to undertake a realistic integrated study of face recognition that takes account of both human faces and human performance are not firmly in place. This study represents only a beginning to attempt to consider the diversity and complexity of this problem.

Acknowledgements

This work was supported by an Alexander von Humboldt Stifung and NIMH grant. 1R29MH5176501A1 to AJO'T. Thanks are due to Niko Troje and Isabelle Bülthoff for the stimulus creation and processing and to K.A. Deffenbacher, J. Liter, D. Valentin and two anonymous reviewers for helpful comments on this manuscript.

References

[1] Light L, Kayra-Stuart F, Hollander S. Recognition memory for typical and unusual faces. J Exp Psychol Hum Learn Mem 1979;5:212–28.
[2] O'Toole AJ, Deffenbacher KA, Valentin D, Abdi H. Structural aspects of face recognition and the other-race effect. Mem Cogn 1994;22:208–24.
[3] Valentine T, Bruce V. The effects of distinctiveness in recognising and classifying faces. Perception 1986;15:525–36.
[4] Vokey J, Read D. Familiarity, memorability, and the effect of typicality on the recognition of faces. Mem Cogn 1992;22:208–24.
[5] O'Toole AJ, Deffenbacher KA, Abdi H, Barlett JC. Stimulating the 'other-race effect' as a problem in perceptual learning. Connect Sci J Neural Comput Artif Intell Cogn Res 1991;3:163–78.
[6] Edelman S. Computational theories of object recognition. Trends Cogn Sci 1997;1:296–304.
[7] Moses Y, Ullman S, Edelman S. Generalization to novel images in upright and inverted faces. Perception 1996;25:443–62.
[8] Troje N, Bülthoff HH. Face recognition under varying pose: the role of texture and shape. Vis Res 1996;36:1761–71.
[9] Troje N, BÏthoff HH. How is bilateral symmetry of human faces used for recognition of novel views. Vis Res 1998;38:79–89.
[10] Lando M, Edelman S. Receptive field spaces and class-based generalization from a single view in face recognition. Network 1995;6:551–76.
[11] Pentland A, Moghaddam B, Starner T. View-based and modular eigenspaces for face recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition, June, 1994.
[12] Valentin D, Abdi H. Can a linear associator recognize faces from new orientations? J Opt Soc Am A 1996;13:717–24.
[13] Vetter T, Poggio T. Linear object classes and image synthesis from a single example image. IEEE Trans Pattern Anal Mach Intell 1997;19(7):733–42.
[14] Duvdevani-Bar S, Edelman S, Howell AJ, Buxton, H. A similarity-based method for generalization of face recognition over pose and expression, In: Akamatsu S, Mase K editors. Proc 3rd Int Symp on Face and Gesture Recognition (FG98). Washington, DC: IEEE Press, 1998.

[15] Ullman S. High Level Vision. Cambridge, MA: Bradford, 1997.
[16] Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. Nature 1990;343:263–6.
[17] Cohen JD, McWhinney B, Flatt M, Provost J. PsyScope: a new graphic interactive environment for designing psychology experiments. Behav Res Methods Instrum Comput 1993;25:257–71.
[18] Valentin D. How come when you turn your head I still know who you are? Evidence from computational simulations and human behaviour. Unpublished PhD dissertation, The University of Texas at Dallas, 1996.
[19] O'Toole AJ, Deffenbacher KA, Valentin D, McKee K, Huff D, Abdi H. The perception of face gender: the role of stimulus structure in recognition and classification. Mem Cogn 1998 (in press).
[20] Edelman S, Duvdevani-Bar S. Similarity, connectionism, and the problem of representation in vision. Neural Comput 1997;9:701–20.
[21] Valentin D, Abdi H, Edelman B. What represents a face: a computational approach for the integration of physiological and psychological face data. Perception 1998;26 (in press).
[22] Rock I, Wheeler D, Tudor L. Can we imagine how objects look from other viewpoints? Cogn Psychol 1989;21:185–210.
[23] Edelman S, Reisfield D, Yeshurun Y. Learning to recognize faces from examples. In: Sandini G editor. Proc 2nd Eur Conf on Computer Vision, Lecture Notes in Computer Science, 588. Berlin: Springerm, 1992, p. 787–791.