Viewpoint generalization in face recognition:

The role of category-specific processes

Shimon Edelman

The University of Sussex at Brighton

Alice J. O'Toole

The University of Texas at Dallas

8, January 1999

Abstract

The statistical structure of a class of objects such as human faces can be exploited to recognize familiar faces from novel viewpoints and under variable illumination conditions. We present computational and psychophysical data concerning the extent to which *class-based learning* transfers or generalizes within the class of faces. We first examine the computational prerequisite for generalization across views of novel faces, namely, the similarity of different faces to each other. We next describe two computational models which exploit the similarity structure of the class of faces. The performance of these models constrains hypotheses about the nature of face representation in human vision, and supports the notion that human face processing operates in a class-based fashion. Finally, we relate the computational data to well-established findings in the human memory literature concerning the relationship between the typicality and recognizability of faces.

Viewpoint generalization in face recognition: The role of category-specific processes

Computational theories of visual recognition that postulate image-based, view-dependent object representations have been gaining both computational and psychophysical support in recent years. These provide an alternative to theoretical perspectives on object representation that assume three-dimensional shape reconstructions (Ullman, 1996; Edelman, 1997). This theoretical development has created a certain tension within the psychological literature on object and face recognition. Specifically, although psychologists consistently find that the human visual system is capable of making sense of an image of an object, even when the object is encountered under novel viewing conditions, it is not immediately clear how a view-based representation can support this ability:

> "What can possibly be left if no parts can be found, if no depth relations can be determined ...? Quite likely, the only remaining process is the simplest of all pattern operations: viewpoint-specific recognition. This is, undeniably, a modern code word for 2-D templates. As would be expected for such a top-down process, it only works for familiar objects, seen from familiar viewpoints." (Cavanagh, 1995).

We attempt to ease the tensions between theory and experiments by showing (1) that given prior experience with objects of similar shape, multiple-view models can be made to exhibit a considerable degree of generalization, both to novel views and to novel objects, and (2) that such models are relevant for understanding human generalization performance on novel stimuli, which, likewise, depends on prior exposure to objects that belong to same shape class.

To characterize the role of learning in generalization across viewing conditions, one needs a class of stimuli that combines relative uniformity with a nontrivial statistical structure. Human faces are uniquely suitable in this respect for a number of reasons. First, faces are a structurally tight class, defined by a set of features (e.g., eyes, nose, mouth, etc) arranged in a particular, universally recognizable configuration; the human race contains billions of variations on this configural theme.[1] Second, despite the highly constrained geometry of faces, people exhibit an amazing ability to discriminate and remember large numbers of individual faces, even across decades (Bahrick, Bahrick & Wittlinger, 1975) – an ability not required normally for other categories of objects such as chairs or cows. Third, the human face processing mechanism is quite flexible in generalizing from a single view of a novel face, tolerating transformations that result in large changes in the 2D image, including changes in illumination and viewpoint, as well as changes in facial expression, hair style, etc.. Finally, an extensive body of psychological research on face perception is readily available, and can be put to use in constraining computational theories of generalization in face recognition. A major goal of this chapter is to make use of converging psychological and computational data for understanding human performance on face processing tasks. The modeling supplies a concrete quantification of the information in faces and can be used to implement arbitrary transformations of this information which can be used to solve psychologically relevant tasks. The psychological data constrain the plausibility of these models as hypotheses about human face processing.

[1] It should be noted that the concept of a facial feature is itself quite controversial. The chapter by O'Toole et al., this volume, considers this controversy in some detail.

The impressive abilities of human observers in face recognition are in part the result of the large amount of experience we have with faces. Face recognition has been considered as an example of a task requiring perceptual expertise (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Rhodes & McLean, 1990). The relevance of expertise as a theoretical construct for describing human face recognition is perhaps best illustrated by a recent study by Gauthier and Tarr (1997). They showed that many of the effects typical of face recognition can be obtained with this computer-generated shapes that parallel faces in their complexity, structural uniformity, and statistical diversity, if the observers are given the chance to accumulate sufficient expertise in a few thousands of trials with this object class (Gauthier & Tarr, 1997).

Other evidence for the importance of expertise comes from developmental work indicating that the ability to remember new faces improves gradually until the age of 12 (Carey, Diamond & Woods, 1980). This learning does not extend automatically even to sub-categories of faces with which we may have less experience, e.g., faces of people of another race (Brigham, 1986). Further, the learning does not extend to inverted faces, which are comparable in complexity, contrast, and spectral content to upright faces (Yin, 1969; Moses, Ullman & Edelman, 1996).

The central unifying thesis of the present paper is the hypothesis that flexibility of human recognition skills with faces results from the specialized, class-based nature of the processes employed by the visual system. By "class-based" we mean processes that are derived from, and hence specific to, a class of objects; faces, in this case. Specifically, we argue that (1) the class-based transformations that operate in human face processing are learned though extensive experience with faces; (2) these learn-by-example processes exploit the statistical structure of the set of faces on which they are based and can be studied, therefore, by measuring their success or failure with individual faces that are either typical or atypical examples of this experience; (3) the flexibility of the transfer of learning between different viewing conditions in face recognition is limited to stimulus objects that the visual system "treats as faces". Thus, while the generalization of recognition within the category of "faces" is quite good, due to the relatively stable statistics of the structurally tight class of faces, outside of this class, the performance of these processes is rather less impressive.

The hypothesis of class-based processes stands in opposition to an implicit view in the computational vision literature that assumes that general visual processes, (e.g., the discounting of an illuminant, viewpoint and orientation normalization) are equally effective and efficient for all objects. We wish to note clearly that class-based processing does not necessarily suggest categorical behavior in the human ability to generalize task performance across these basic visual processing challenges. Rather, the hypothesis states that our generalization ability should be a function of the degree to which the class-based statistics apply to the stimulus in question. This clearly is a function both of the nature of the face representation, and, as we shall see shortly, the task.

Given our emphasis in this chapter on the category of human faces, we wish to be clear about our claims concerning the generality of the model we propose. Much effort has been expended in the literature in recent years to determine whether or not faces are "special". We believe the use of the word "special" to describe face processing has lead to quite a bit of unnessessary confusion in the literature as to what precisely is meant by "special". Notwithstanding, there is actually very little disagreement in the literature that faces, as a category of objects, are "special" in several ways. Farah, Wilson, Drain, and Tanaka (**?**) provide an excellent review of these factors, and we shall touch on nearly all of those they mention in the course of the present chapter. Given these factors, a truely "general" computational model of face recognition may be neither possible nor desirable as a good model for human face recognition. Indeed we will argue that the human visual perceptual system accomplishes its impressive range of tasks with faces using special-purpose computations that will not generalize well to other categories of objects. They will not generalize well to other objects because they depend on the possibility of exploiting the statistics of the known and well-experienced class of faces.

In fact, we will argue there is little need to generalize the tasks we accomplish for faces to other non-expert subordinate categories, because there is little psychological evidence that they are relevant. Perhaps the most distinguished aspect of human face processing is our ability to identify very large numbers of individual exemplars. We challenge the reader to generate other categories of objects for which we must remember equally large numbers of individual exemplars (e.g. chairs, coats, suitcases, cars?). Consider how many exemplars from these categories must we encode and remember as individuals. The answer to this question goes a long way in understanding why faces are so different from other categories of objects.

Throughout this chapter, we will point out the contrast between the power of special purpose models that operate optimally within the category of *faces* and the limitations of these models for operating beyond the category of faces. Despite the face-specific nature of the model we propose, in several other respects our model has quite general implications for problems requiring perceptual learning and expertise (Gauthier & Tarr, 1997), and for understanding mechanisms that can implement complex categorizations in a multidimensional space (?). Perceptual learning, which we use in the present model, is a theoretical construct relevant for understanding expertise across a broad range of perceptual and cognitive tasks. Such a mechanism has been proposed for understanding human expertise in processing faces in much the same way as it is thought to underlie other domains of perceptual expertise (O'Toole, Abdi, Deffenbacher & Valentin, 1995). Further, models of categorization in multidimensional spaces are a staple of many general mathematical approaches to understanding representation (Townsend & Thomas, 1993). To these approaches, the present work adds a *physical* instantiation of a face representation that reflects the perceptual richness and complexity of the stimulus. Townsend and Thomas (1993) and others have argued that the general natural of abstract models of categorization are limited in the extent to which they include the natural structure of the stimulus, in this case, the rich perceptual stucture of faces, in structuring the process of recognition and categorization.

In the following sections, we present computational and psychophysical data concerning the extent to which class-based learning generalizes within the class of faces. To address this question, we first consider the somewhat surprising psychological "boundaries" of the class of human faces for human perceivers. We describe a human face recognition experiment with upright and inverted (i.e., upside-down) faces. Despite the fact that the contrast, feature complexity, and similarity structure of inverted and upright faces are identical, the performance of the human subjects in generalizing recognition across changes in viewpoint is very poor for inverted faces relative to upright faces. A knowledge of the psychological boundaries of the category of faces constrains hypotheses about the nature of representations humans employ for faces. Using these constraints, we next describe a computational model of face recognition that makes use of a normalizing class-based procedure for recognizing faces across changes in viewpoint. Computational data on the statistical reliability of the transformations for exemplar faces are presented, comparing different kinds of representations. Finally, we focus on factors that affect performance within the class of faces by examining the role of face typicality/distinctiveness in the success of class-based processes.

In what follows, the reader will find connections to a number of concepts that are considered elsewhere in this volume. These concepts include face space representations, distinctiveness, recognition over changes in viewpoint, and morphing. Especially related are the chapters by Townsend, Solomon and Spencer-Smith and Valentine, which consider the properties of face space representations. Additionally, in the chapter by Busey, questions about face distinctiveness are addressed by using morphing techniques to navigate through a simulated face space. The importance of face distinctiveness is also considered in the chapter by Valentine. The problem of recognizing faces across changes in viewpoint is treated thoroughly in the chapter by Valentin, Abdi, Edelman, and Posemetier. Finally, O'Toole, Wenger and Townsend cover basic issues concerning the nature of underlying representations that are relevant for putting the present model into the broader context of computational models of face processing.

Evidence for class-based processing

Human performance in recognizing faces has been studied psychophysically for many years. The most intriguing results, and those that have yielded the most useful information about how we represent and recognize faces, have come, surprisingly, from the tasks on which human face recognition fails, a fact that may be due to our extraordinary abilities with faces over other objects. For example, it is well known that human observers have great difficulty in recognizing upside down faces relative to other objects (Yin, 1969). A rather striking illustration of the problems we experience in perceiving upside-down faces can be seen in the “Thatcher illusion” (Thompson, 1980), so named because it was first demonstrated with Margaret Thatcher’s face (Figure 1). This illusion shows that we are remarkably insensitive to even gross distortions in a face when viewing it upside down.

Bartlett and Searcy (1993**?**) have employed the Thatcher illusion as a tool for asking questions about the effects of inversion on processing configural versus componential information in faces. They found that inversion impairs the processing of configural information in faces. The relative emphasis on configural or holistic encoding of faces has been further supported by Farah et al. (**?**). An very interesting component of the role of holistic processing in the inversion effect, however, it that it appears to be bound to the kind of task being performed. Searcy and Bartlett (**?**) found, for example, no difference in response rates for a same-different task in for pairs of upright versus faces when the faces were normal, or were altered in a componential rather than configural way. In general, they conclude that tasks encouraging componential processing of faces are less affected by inversion than those encouraging face processing (Bartlett & Searcy, 1993**?**).

The difficulty we have in processing inverted faces, seen so clearly in the Thatcher il-

Figure 1: A “Thatcherized” face. Turn the page upside down to see the full extent of the face distortion

lusion, is but one of several classic illustrations that converge on the importance of role of expertise in processing the configural information in faces. Other examples include the processing of faces in the photographic negative (Galper & Hochberg, 1971), the processing of “other-race” faces (e.g. Brigham, 1986), and the processing of faces with unusal illumination (Hill & Bruce, 1994, 1996).

As noted previously, the class-based hypothesis states that generalization ability should be a function of the degree to which the class-based statistics apply to the stimulus and task in question. The expertise results just considered indicate that the nature of the face representation on which these statistics are built is likely to be mono-oriented and configural in nature. Thus, we expect that our general, universal, visual processing procedures will generalize best to stimuli that lie within our domain of expertise.

A Psychophysical Delineation of the Class of Faces

Evidence for the operation of class-based processes with human faces can be obtained by examining the extent to which the human visual system can generalize recognition across changes in illumination and pose or viewing position of faces. Illumination and viewpoint changes affect the appearance of the image of an object. Processing variation in illumination and viewpoint is a basic visual pre-requisite tasks for scene processing, and navigation — not just for object categorization and recognition. These generalization tasks within the domain of faces have been studied recently by Moses, Ullman and Edelman (1996). They tested human subjects in a forced-choice discrimination task with either upright or inverted faces (Moses, Ullman & Edelman, 1996). We use this study to illustrate that while class-based generalization procedures operate efficiently and accurately on upright faces, they operate much less efficiently and accurately for a stimulus that is equally complex and statistically similar (spectral content, contrast, etc.), but is not a face.

Observers in the experiments of Moses et al. (1996) were tested in a series of 24 sessions. In each session, observers first learned to discriminate among images of three faces, taken under a fixed viewing position and illumination. In half of the sessions they discriminated among three upright faces; in the other half of sessions they discriminated among three inverted faces. Generalization *within* each of these two families of stimuli was tested using images of the same faces taken under all combinations of four illuminations (left, center, right, and combined) and five viewing positions ($-34°$, $-17°$, $0°$, $17°$, and $34°$ from the frontal view in the horizontal plane, controlled by a moving robot arm).

For upright faces, near-perfect recognition and equally good generalization to novel conditions were found. For inverted faces, although recognition performance after training was similar to the performance for upright faces, the generalization to novel views was significantly worse for inverted compared to upright faces. The generalization in the inverted condition improved with practice, even in the absence of feedback. After four sessions with the same stimuli, performance improved for the novel illumination and view position conditions. Notably, however, this improvement *did not transfer* to novel inverted faces, with which the observer had no prior experience.

These results provide evidence for two kinds of constraints operating in human generalization processes for faces. First, because observers did not generalize as easily for inverted faces as for upright ones, it is clear that at least some of the processes supporting generalization across viewpoint and illumination are not universal. Second, the nearly perfect generalization found for upright faces from only a single view, by itself insufficient for building a complete object-specific model, indicates that the generalization processes cannot be strictly object-specific. These results led Moses et al. to propose that generalization in face recognition occurs at an intermediate computational level that is applicable to a *class* of objects — and that at this level upright and inverted faces initially constitute distinct classes of objects.

In the context of class-based processes, the key factors that cause inversion effects for faces are the large differential in experience humans have with upright versus inverted faces and the structurally homogeneous yet statistically complex nature of face geometry (Rock, 1973, 1988). The combination of these two factors is typical in many ways of several classical cases of perceptual learning (O'Toole, Abdi, Deffenbacher & Valentin, 1995). As in those cases, there is need to deal with the highly complex nature of the information in faces. Because faces are usually seen in an upright orientation, orientation-specific processes may be a reasonable design choice for a biological visual system which must allocate its limited resources in a judicial manner depending on the statistics of the stimuli with which it is confronted. Note again, that such effects can be found for other objects when similar constraints apply (Diamond & Carey, 1986; Gauthier & Tarr, 1997).

Quantifying Face Similarity: A Physical Substrate for Class-based Processing

Extensive experience, which, by hypothesis, constitutes the computational basis for the remarkable ability of human observers to generalize from single views of upright faces, can be put to use under one condition. Namely, faces must be sufficiently similar to each other to allow the visual system to estimate the effect of a given view transformation on the appearance of a novel face. In the remainder of this section, we examine the extent to which this condition is satisfied for human faces and for computer-generated face-like objects.

Representing Face Images

Before attempting to quantify face similarity, one must decide upon the proper manner of representing face images for this purpose. Raw pixel images might be a poor choice, given what is known about early visual processing of images. To emulate to some extent the initial stages of biological visual processing, face images can be represented by *vectors* of activities of graded overlapping receptive fields, similar to those used in most models of early visual processing following Marr (1982). An additional advantage of this mode of representation is its relative insensitivity to certain changes in the viewing conditions. On the one hand, raw pixel images of the same face taken under different viewing conditions are likely to be less similar to each other, under a reasonable choice of image metrics, than images of different faces taken under identical conditions (Moses, Adini, & Ullman, 1994). On the

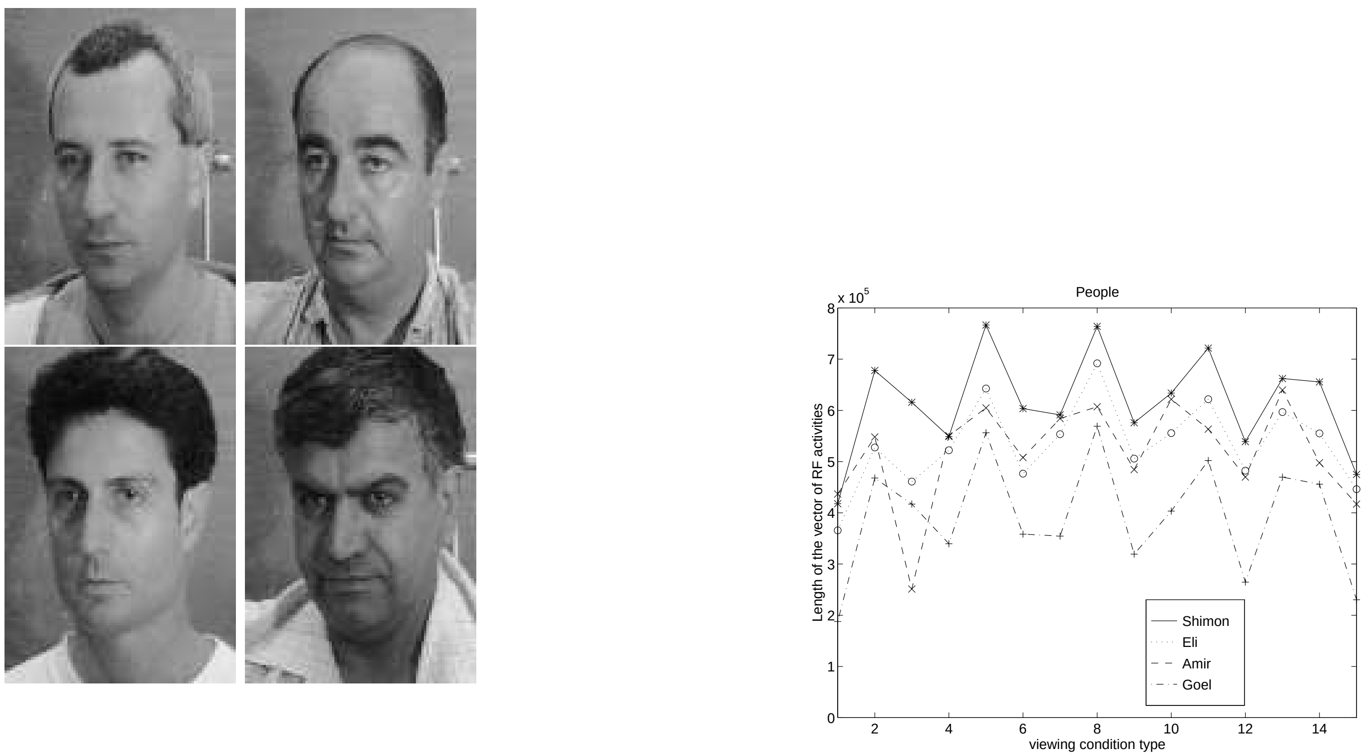


Figure 2: Left: four human faces from the database of Moses, Ullman and Edelman (1996). Right: length changes of the RF-space vectors corresponding to the four human faces, over 15 different viewing conditions. Note that the changes for different faces are correlated with each other (see Lando and Edelman, 1995b).

other hand, using large overlapping RFs with a bandpass spatial frequency response curve instead of pixels can alleviate this problem to a considerable degree (Weiss & Edelman, 1995). Although we do not consider this a complete or exhaustive model of early visual processing of images, we have tried to exploit the most important advantages of this code over the use of raw image data.

Human faces

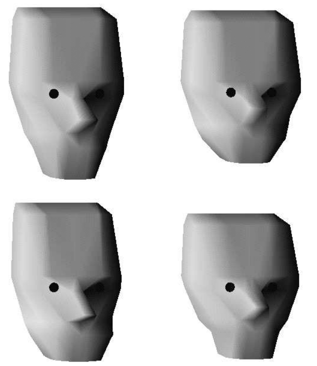
We now turn to the question of the influence of changes in the viewing conditions upon the receptive field representations of human faces (see also Valentin, et al., this volume).

First, however, we note that the vector-based representation of faces can be used to define a face space (O'Toole, Abdi, Deffenbacher & Valentin, 1993, 1995) with the dimensions/axes corresponding to the receptive field samples from the images. The notion of a *face space* was introduced by (Valentine, 1991) as an abstract psychological model of face representations. More concretely, face space representations underlie most computational models of face recognition and categorization. For present purposes, a face space underlies the current model as follows. A face is represented by a point in a high dimensional receptive field space. A change in the viewing conditions results in a shift of the point to a new location in the receptive field space.

The possibility of viewpoint and/or illumination generalization depends on the relative predictability of the point shifts for different faces as a function of specific changes in the viewpoint and/or illumination conditions. Note that this requires predictability both in the direction and the extent of the shift. A study by Lando and Edelman (Lando & Edelman, 1995) involving the same human face database used in (Moses, Ullman & Edelman, 1996) showed that the directions of the shifts for various faces match each other to within a few degrees; the extent (length) of the shift is similarly correlated for various faces, as illustrated in Figure 2. The predictability of the direction and extent of point shifts in this space comes from the highly similar structure of human faces. Such predictability is not a property of objects in general, but rather holds only within a class of objects, such as faces.

Model heads

In addition to giving a statistical characterization of the homogeneity of the receptive field space transformations of human faces, as induced by view changes, the study of Lando



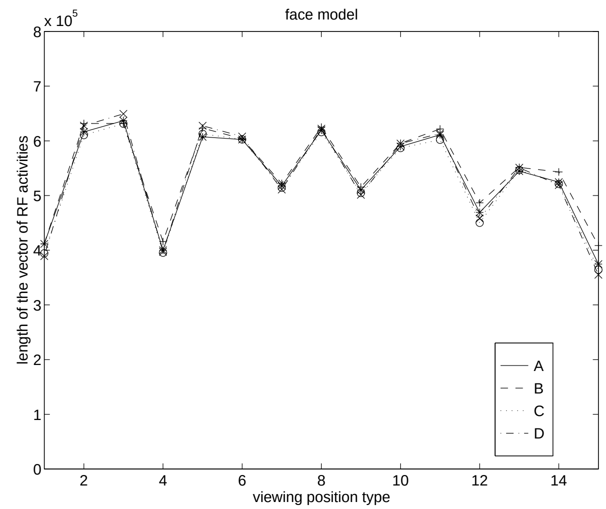


Figure 3: Left: four computer-generated stylized models of human faces. Right: length changes of the RF-space vectors corresponding to four different computer-generated face models, over 15 different viewing conditions. Note that the changes for different faces are correlated with each other (see Lando & Edelman, 1995b).

and Edelman also derived quantitative bounds on deviation from homogeneity. This was done by creating a three-dimensional computer model of a stylized face composed of several dozen triangular facets. Two-dimensional images of these faces were rendered using standard illumination and surface reflectance models. The parameters of the stylized triangular mesh were then varied to model the natural variability in human face geometry (see Figure 3). It was then possible to derive bounds on the angle and length dispersion of the receptive field space representation shifts for these modeled faces. The analytically derived bounds for the human faces and those derived from the results of computer graphics simulations were shown to be comparable and well within the theoretical bounds (Lando & Edelman, 1995).

A Viewspace-interpolation Model of Class-based Processing

It is possible to capitalize on the orderly behavior of receptive field space representations of faces under view changes in replicating the class-specific generalization effects found in the psychophysical studies. In this section, we show how a multiple-view model of recognition can be adapted directly to carry out class-based generalization; a more elaborate model is described in the Model Section.

Viewspaces

Consider a multidimensional space of measurements (e.g., the activities of receptive fields) performed by a visual system upon the world. A view of a face corresponds to a single point in this space; a smoothly changing scene such as a sequence of views of a face rotating in front of the observer — to a smooth manifold that we may call the *viewspace* of the object. The dimensionality of the viewspace depends on the number of degrees of freedom of the object; a rigid object rotating around a fixed axis gives rise to a one-dimensional viewspace. This can be seen in Figure 4, where we see two viewpoint disparate images a particular individual, connected by the curved trajectory $\mathcal{V}_1$.

By continuity, the viewspaces of two nearly identical faces will be very close to each other; a smooth morphing of one face into another will result in a concomitant smooth evolution of its viewspace, if the measurement functions are themselves smooth. This observation can be turned into a computational basis for the treatment of novel objects (Edelman & Duvdevani-Bar, 1997a), and, in particular, novel faces. Suppose that a system has internalized the viewspaces of a number of faces; it can then process a novel view of a novel face intelligently, to the extent that it resembles the familiar faces (see Figure 4). For this to work, the concept of similarity must be given a concrete interpretation in terms of the measurement space. A computational mechanism suitable for this purpose is interpolation, with the important grounding of reference or prototype faces. These reference faces ground the space in a way that provides general information about the result of common tranformations of view and illumination. Thus, the change in the view (appearance) of a person unfamiliar to the system, (i.e., previously seen from only one viewpoint), can be estimated by interpolating corresponding changes in the appearance of reference (prototype) faces which have been seen before from many viewpoints.

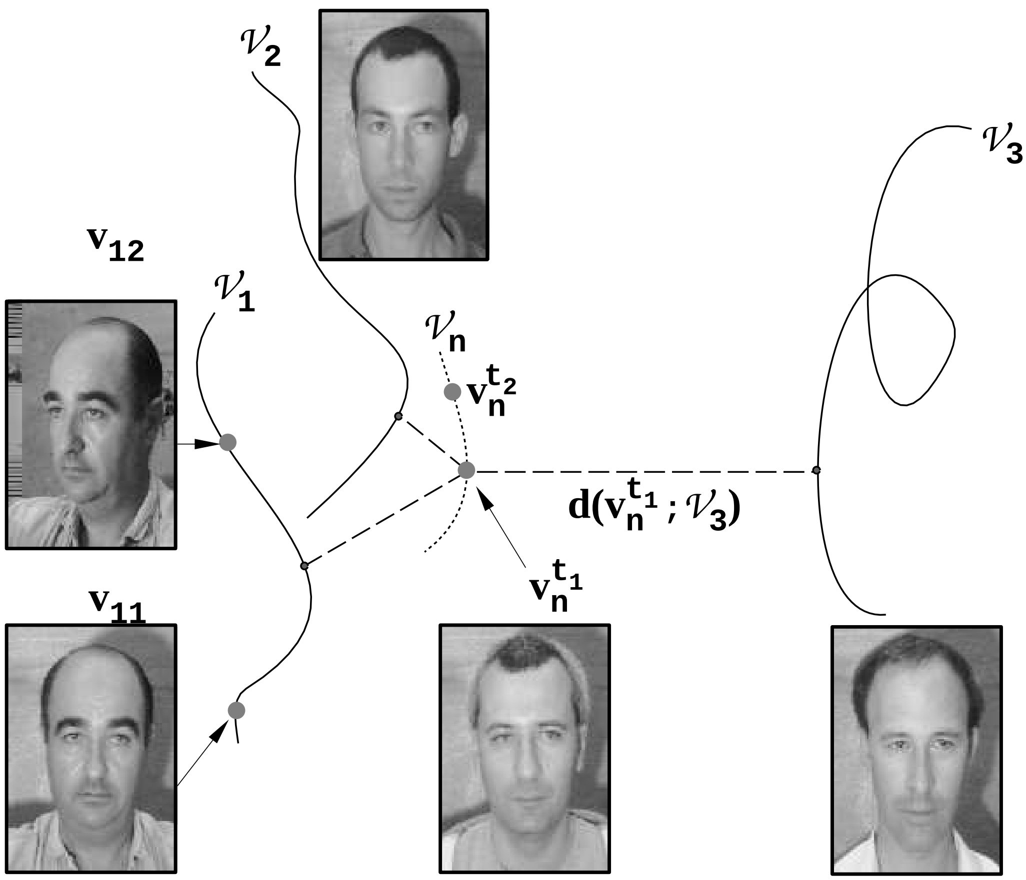


Figure 4: Interpolation of prototypical viewspaces, after Edelman and Duvdevani-Bar (1997). The change in the view (appearance) of a person unfamiliar to the system, (i.e., previously seen from only one viewpoint), can be estimated by interpolating corresponding changes in the appearance of reference (prototype) faces which have been seen before from many viewpoints.

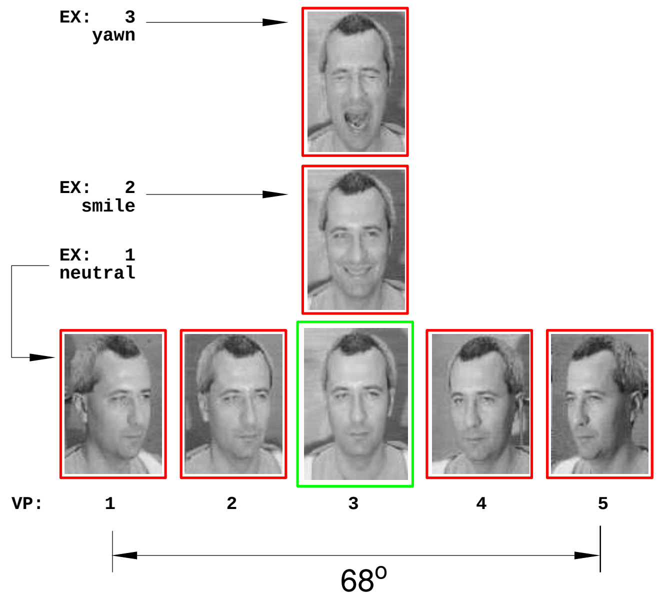


Figure 5: The dimensions of variation in the face data used in testing the viewspace interpolation model. For each of the 28 persons included in the database, which was a subset of the Weizmann FaceBase Moses, Ullman, and Edelman (1996), 15 face images were used (corresponding to 5 viewing positions in increments of 17° $\times$ 3 expressions). The face with viewpoint (VP) = 3, expression (EX) = 1 was used as the single image from which generalization was tested.

Direct Interpolation of Viewspaces

The interpolation of viewspaces involves irregularly spaced data, because the distances among the viewspaces (i.e., the similarities among the faces) need not conform to any regular *a priori* pattern. Among the many interpolation methods that can treat irregularly spaced data (Alfeld, 1989), inverse-distance weighting (Shepard, 1968; Gordon & Wixom, 1978) seems to be the simplest. In this algorithm, the contribution of a known data point to the interpolated value at the test point is inversely proportional to the distance between the two. In terms of Figure 4, we wish to compute the shape of the viewspace of a new face by interpolating among the shapes of the viewspaces of familiar faces. Thus, our data "points" are actually entire manifolds – the viewspaces of the reference faces. Accordingly, the success of the interpolation approach here depends on the availability of a mechanism for dealing with entire viewspaces of individual familiar faces. An example of such a mechanism is the Radial Basis Function (RBF) module, which can be trained to output a constant for different views of its target object (Poggio & Edelman, 1990).[2] Because its output decreases monotonically with the dissimilarity (i.e., distance) of the stimulus from the viewspace of the object on which the module had been trained (Edelman & Duvdevani-Bar, 1997b), it is precisely what one needs for the inverse-distance weighted interpolation. Specifically, consider a system composed of k modules, each trained to output 1 for a number of representative views of some reference object. As observed above, the output of the i-th module for a given test view $\mathbf{v}_n^t$ of a novel object, $x_i(\mathbf{v}_n^t)$, can serve as an indicator of the relevance of the i-th prototypical viewspace $\mathcal{V}_i$ to estimating the structure of the viewspace of the novel object $\mathcal{V}_n$. Consequently, the weight of $\mathcal{V}_i$ in determining the shape of $\mathcal{V}_n$ should be set to $x_i(\mathbf{v}_n^t)$.

One way to apply this principle to recognition is to compute a quantity Y that would remain constant over changes in the test view $\mathbf{v}_n^t$ of a novel object (Edelman & Duvdevani-Bar, 1997a). Let the vector of responses of the k modules to a test view t_1 be $\mathbf{w} = \mathbf{x}(\mathbf{v}_n^{t_1})$. The estimate of Y for another test view t_2 is then $Y\left(\mathbf{v}_n^{t_2}\right) = \mathbf{w}^T \mathbf{x}\left(\mathbf{v}_n^{t_2}\right)$, where T denotes the transpose. Note that the weights are pre-computed for a certain input, then used for other inputs (i.e., in other parts of the input space). $Y(\mathbf{v}_n^{t_2})$ will remain approximately constant, as long as the test view $\mathbf{v}_n^{t_2}$ is not too far from the view $v_n^{t_1}$ used to estimate the weights $\mathbf{w}$, and as long as the novel object is not too different from at least some of the reference ones.

[2]Psychophysical and physiological data (Bülthoff & Edelman, 1992; Logothetis & Pauls, 1995; Logothetis, Pauls & Poggio, 1995) suggest that this interpolation mechanism is particularly relevant to the modeling of recognition in biological visual systems.

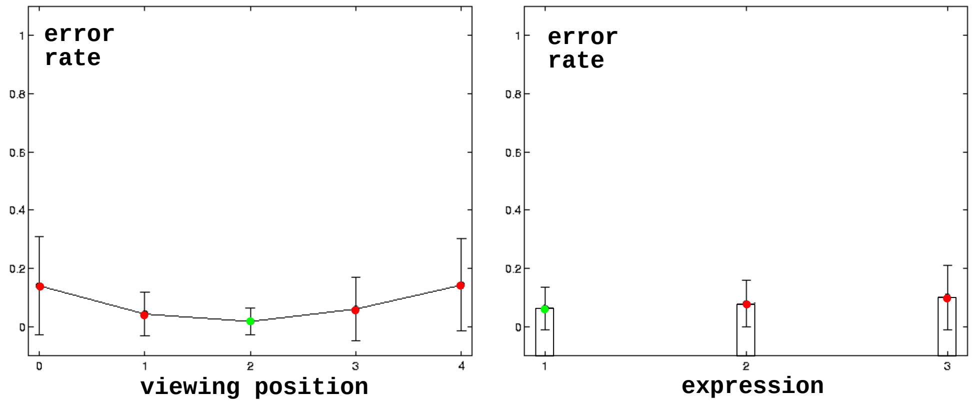


Figure 6: Performance of the system in the 3-way classification task. *Left:* error rate vs. viewpoint, averaged over the three different values of expression. *Right:* error rate vs. expression, averaged over the five different values of viewpoint. The mean error rate over the five viewing positions and the three expressions was 0.08. Human subjects in a comparable task exhibited an error rate of about 0.03 Moses, et al. (1996).

Tests of this approach were conducted on images of faces that differed along two dimensions: orientation (rotation of the head around the vertical axis) and expression. A subset of the images from the 28-person Weizmann FaceBase (Moses, Ullman & Edelman, 1996) was used in the experiments (Figure 5). Ten of these faces were used as training faces, which formed the reference face modules for the simulation. The remaining 18 faces were reserved for testing purposed. All images were cropped and subsampled to a size of 100×100, then convolved with a bank of Gabor filters (Howell & Buxton, 1995). The filters were at four scales and three orientations, and formed a sparse, non-overlapping grid to provide 510 coefficients per image. To train the 10 face modules, 15 images, corresponding to all the combinations of five orientations and three expressions of each of the 10 training faces were employed. A reference face module was thus defined for each of these 10 faces. Similarly prepared images of the remaining 18 faces (270 images altogether) were used to test the generalization ability of the system.

For each of the 18 test faces, the image corresponding to a full face orientation and neutral expression was used as the single view from which generalization was to be carried out. The vectors of 10 module responses to that single view of each of the 18 test faces were pre-computed and used as the sets of weights in the generalization test stage. During testing, the system computed the weighted sum of the 10 module responses using each of the 18 sets of weights in turn. The set that yielded the highest sum out of the 18 possibilities determined the identity of the test view. The mean error rate in this 18-way recognition task was about 31%. To facilitate the direct comparison of this system's performance to that of human observers, a 3-way discrimination experiment was conducted, which paralleled the setup of Moses et al. (1996). In that study, subjects carried out 3-way discrimination of faces drawn from the same database used here, achieving about 3% error rate for generalization over viewing position. In comparison, the mean error rate exhibited by the viewspace interpolation system (over the 816 triplets, or all possible combinations of three out of 18 faces) was about 8% (Figure 6), for generalization over viewing position *and* expression — a level of performance that is rather encouraging, albeit still below that of human subjects.

A Two-stage Model of Class-based Processing

The second model we outline relies on the statistics of a collection of face shapes in two ways. The common manner in which images of faces change with viewpoint, due to the common 3D structure of faces, is exploited at the initial stage of the model, which performs normalization of the input image to a "standard" view of the face. The normalized image is then compared to a number of reference faces, which span the face space for this model. In terms of the illustration of the face space that appears in Figure 4, the first stage of this model "collapses" the viewspace of the stimulus to a point, while the second stage computes the distance of this point to a number of "landmarks" (i.e., collapsed viewspaces of reference faces).

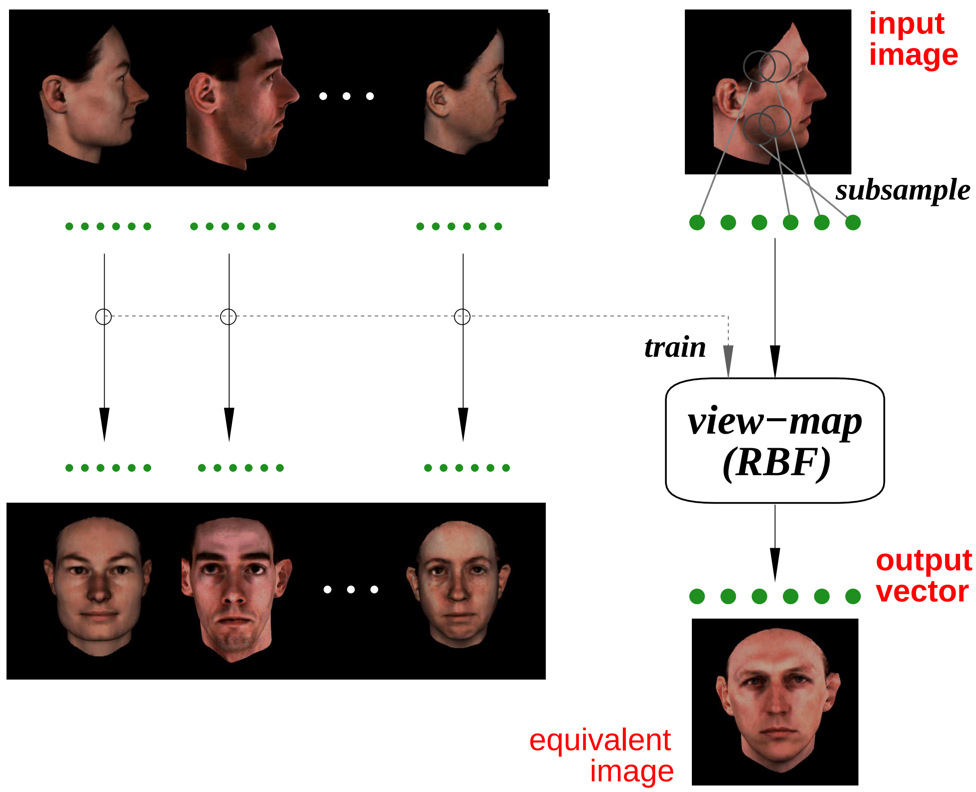


Figure 7: The view-mapper. The way in which known faces change across viewpoint is exploited in deriving a normalized representation of a novel face seen from a familiar orientation.

Normalization in the View Space

A normalizing transformation that brings familiar members of the class of faces into a normal form can be used to estimate the appearance of a less familiar face from some standard viewpoint, making possible subsequent recognition of that face, even if it has been seen before from a single viewpoint. The function of this "view-mapper" fits into the idea we presented previously about interpolating faces in view spaces, but is implemented more efficiently in this model. The early version of the model (Lando & Edelman, 1995) used the mean transformation of the familiar faces to transform the novel one; the present version learns to estimate an optimal transformation from examples, using a linear association network, which acts as a view mapper (Figure 7). This provides a better estimate of the view transformation for individual faces than can be obtained by applying an average transformation.

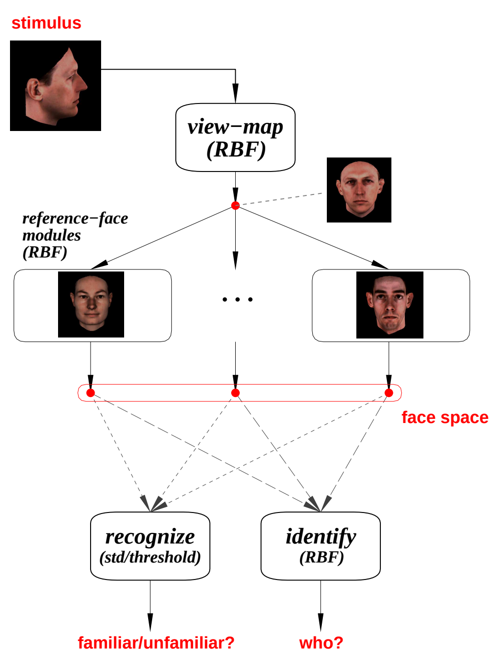


Figure 8: The entire model. Following normalization of the stimulus image by the view-mapper following Lando and Edelman (1995b), the face is projected into a view-specific face space spanned by a set of reference faces, c.f., Edelman, Reisfeld, and Yeshurun (1992).

Localization in the Face Space

At the recognition stage, the system must deal with a face that may have been properly normalized (by a class-based viewpoint transformation), but may still turn out to be unfamiliar, i.e., may not match any of the stimuli for which internal representations are available in long-term memory. Now, observe that a novel face can be represented by locating its corresponding point in the face space. This can be done by estimating the proximity of the face to a number of reference faces, i.e., the similarity of the stimulus face to a number of reference faces (Edelman, 1995b; Edelman & Duvdevani-Bar, 1997b). Thus, the representation of the novel face is via a vector of distances to the reference faces. As we saw in section **??**, this can be done by a simple and biologically relevant mechanism — a radial basis function (RBF) networks (Poggio & Girosi, 1990). In this case, a number of RBF modules are trained, each on images of a different reference face, as illustrated in Figure 8.

Effects of Face Distinctiveness

We now set out to explore the performance of the two-stage model of class-based processing as a function of the statistical structure of the face space. Note that we expect the model to perform differently for typical vs. distinctive faces, because of its heavy reliance on the the similarity structure of the training face set. The aim of this section, which summarizes the results first reported in (O'Toole & Edelman, 1996), is to quantify the accord between the effects of face distinctiveness on the model performance and its effects on human performance, as described in the literature.

Even within the subcategory of faces with which we have the most experience, individual faces obviously vary in the quality of the uniqueness information they provide for a face recognizer — either human or computational. Clearly, the problem of face recognition requires the ability to represent the information in individual faces that makes them unique or different from all other faces in the world. Generalizing across changes in viewpoint entails the additional requirement that this unique information be accessible across viewpoint variations. A face that is unusual or distinct will be less likely to be confused with another face. Indeed, one of the most reliable findings in the psychological face memory literature is that distinctive or unusual faces are more accurately recognized than are typical faces (Light, Kayra-Stuart & Hollander, 1979; O'Toole, Deffenbacher, Valentin & Abdi, 1994; Valentine & Bruce, 1986).

The computational model we proposed in the preceding section applies a class-based transformation to align individual exemplars to a normalized form, then codes the normalized exemplars by their similarities to other exemplars. The performance of this model for individual faces will clearly depend on the extent to which faces are typical or unusual with respect to our experience – though this dependency operates in an interesting and paradoxical fashion at different stages in the process. When the recognition decision does not require normalization to a learned view, typical faces should be recognized more accurately than unusual faces. When class-based transformations are required for recognizing a face (because the face has never been seen from that viewpoint), however, the story becomes more complicated. Because the view mapper is trained with a set of example faces, the quality of normalized view estimates (i.e., how similar the view estimate is to the actual face image from the learned view) should be better for faces that are typical rather than unusual with respect to this transformation. In other words, the view-mapping procedure succeeds insofar as the face is close to the average face, i.e., is typical, and can be approximated in the new view with general information extracted from a set of learned faces.

The two stages of the computational model set up a somewhat paradoxical situation.

Typical faces, which are likely to be the most accurately view-mapped, are not necessarily expected to be the easiest to recognize. This is because typical faces, once view-mapped, are likely to be more similar to, and hence confusable with, other faces than are unusual faces. This 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 in the learning and reference sets. Thus, it becomes possible that relatively unsuccessful view-maps (e.g., for an unusual face), do not necessarily lead to poor recognition, since the face, even badly approximated, may have few similar competitors vis a vis the similarity structure coding (Newell, Chiroro & Valentine, 1996; O'Toole, Edelman & Bülthoff, 1998).

Distinctiveness Simulations

Our major point is that recognition in the computational model involves a trade-off between the success/failure of the normalization and interpolation processes. To examine this tradeoff in a systematic fashion we carried out two sets of simulations. The first simulation assessed the effects of face distinctiveness on the performance of the normalization procedure, and the second assessed its effects on the quality of the resulting face-space representations.

Stimuli: Parameterized Human Heads

To characterize the effect of face distinctiveness on the functioning of the model, we had (1) to quantify the distinctiveness itself, and (2) to obtain a series of faces varying along the distinctiveness dimension. For this latter purpose, one may use synthetic parametrically

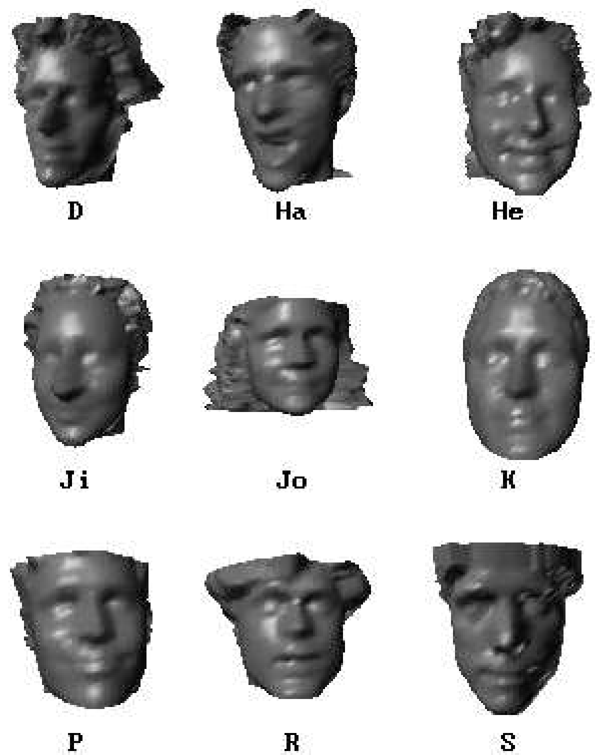


Figure 9: The nine faces used in the generation of the stimulus set.

controlled shapes (Edelman, 1995a; Lando & Edelman, 1995), or derive the parameter space from a set of real faces.

We decided to use of laser scans rather than face images because the laser scans can be rendered at arbitrary viewpoints and under arbitrary illumination conditions. The scans we use are the only ones we know of that are truly in the public domain and are thus accessible to all. Three of these are distributed with SGI systems, and the other six are available over the Internet, courtesy of Cyberware Inc., as a part of their demonstration software).[3] The heads appear in Figure ??. Finally, we note that because we have used only a few scans, we treat the work presented in this paper more as an introduction and exploration

[3]A similar approach to the generation of parametrically controlled face stimuli has been recently proposed in (Atick, Griffin & Redlich, 1996).

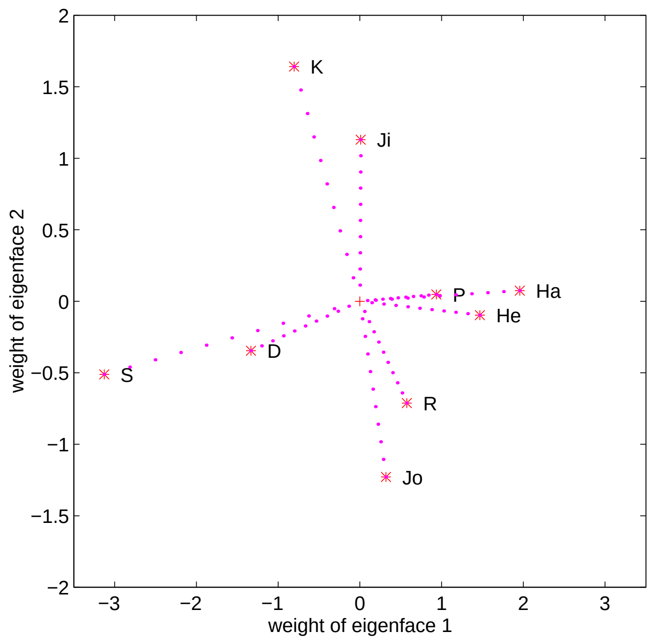


Figure 10: The weights of the nine faces used in the generation of the stimulus set, in the space of the first two eigenheads.

of plausible methods than as indicating generalizable conclusions about the complex nature of face distinctiveness.

Because synthetic faces offer only a crude approximation to the rich 3D structure of the human face, we decided to derive the dimensions of the shape space from a principal component analysis (PCA) of nine 3D laser scans of human faces (see Figure 9). This approach to the parameterization of the face spaces leads to a natural quantification of distinctiveness in terms of the parameter-space distance between a given face and the mean face, the parameters of a face being its projections onto the eigenheads obtained by the PCA.[4] Given the small number of heads, we focused primarily on the gross overriding global features (e.g., head shape) as opposed to more local shape features (e.g., wrinkles).

[4]We use the term "eigenhead" because the PCA operated directly on the 3D head data.

To measure these global variations in shape, we represented the locations of the nine faces used in the PCA in the subspace spanned by the first two eigenheads (O'Toole, Abdi, Deffenbacher & Valentin, 1993). We used eight of the faces[5] to generate 80 face stimuli, in the following manner. For each of the eight points in the face space, 10 versions were generated, corresponding to 10 equally spaced locations along the line connecting that point to the origin. For convenience and for later reference, faces numbered 1, 11, 21, . . . , 71 were the least distinctive versions of the eight faces, while faces 10, 20, . . . , 80 — were the most distinctive versions of these faces. Each of the 80 faces was rendered from four viewpoints, starting at a full-face orientation, and proceeding by rotation around the vertical axis (in 22.5° increments) to 67.5°.

Face Distinctiveness and the View-mapper

Separate linear view-mappers were trained to produce estimates of the full-face view from each of three other views: 22.5°, 45°, and 67.5°. To test the generalization performance of the view-mappers, we employed standard "leave-one-out" cross-validation: a view-mapper was trained with all 10 distinctiveness versions of seven faces and was tested with all 10 distinctiveness versions of the "left-out" face. This procedure was repeated for all eight faces, resulting in view-mapped full-face estimates for all eight faces from each of the three views.

We first assessed the quality of the view-mapped face estimates as a function of face distinctiveness. View-map quality was measured as the cosine of the angle between the original full-face view and the view-mapper's estimate of this view (both defined as vectors). The

[5] Omitting face P, whose direction relative to the origin in the face space nearly coincided with that of Ha.

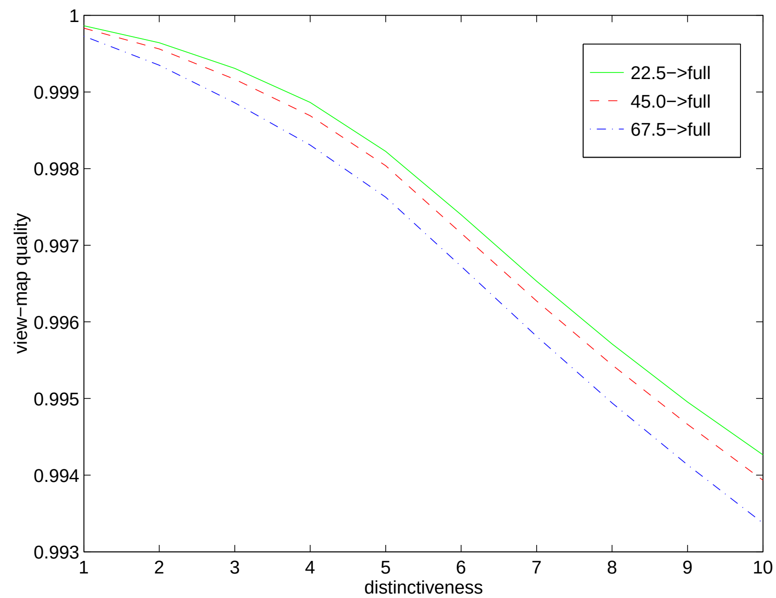


Figure 11: The performance of the view-mapper declines with face distinctiveness and with the disparity between the input and normal views.

results (Figure 11) show that: (1) view-map quality declines as view-map angle increases; and (2) view-map quality declines as the face distinctiveness increases (i.e., typical faces were better preserved than distinct faces in the normalization process, as expected).

Recognition of faces across viewpoint depends not only on the quality of the normalized (view-mapped) face estimate, but also, critically, on the extent to which the structure of face space is preserved across the normalization transformations. We examined this latter issue by analyzing the Procrustes distortion (Borg & Lingoes, 1987) between the original full-face views and their view-mapped versions. This was done by applying Procrustes transformations to compare the similarity of original and view-mapped configurations, in which each face was represented by its coordinates in the space of the two leading eigenvectors derived from the face *images*. The Procrustes analysis determines the optimal combination of scale, rotation, translation, and reflection that minimizes the sum of squared distances between the corresponding points of two configurations. The resultant Procrustes distance is the residual that remains after the application of the optimal transformation, and measures the discrepancy between the two configurations. This distance was 2.91 for the 22.5° view-map condition, 3.18 for the 45° view-map condition, and 4.04 for the 67.5° view-map condition – all significantly better than estimates of the expected random distance, obtained by bootstrap (Efron & Tibshirani, 1993), indicating the preservation of the original similarity structure of the face space by the view mappers.[6]

Finally, we examined the extent to which face distinctiveness influenced the distortion of the face space under view-mapping, by comparing Procrustes distances between the original frontal views and view-mapped versions of the faces for different levels of distinctiveness (see Figure 12). We found that the face space distortion increased with the size of the view change. Moreover, there was a relatively consistent relationship between face-space distortion and distinctiveness, with the lowest distortion for the least and the most distinct faces. Thus, while Figure 11 shows that *view-map quality* declines with increasing distinctiveness, the extent to which the *structure of the similarity space* is preserved does not follow a similar decline. Note that the rise in the distortion with distinctiveness suggests that the view-mapper loses more information from the distinct faces than from the typical faces. There is, however, more uniqueness information in the distinct faces to begin with; this effect, apparently, more than cancels the previous one, resulting in a downward trend

[6]Note that the above analysis was concerned with the preservation of the information in face images, rather than in the 3D head data. Procrustes analysis of the relationship between the similarity space of the 3D head data and that of its 2D representation (a full-face view) indicated that the 3D head and 2D view face spaces did not match well. In other words, view-based and 3D face codes make rather different predictions about the distinctiveness of individual faces; cf. (O'Toole, Vetter, Troje & Bülthoff, 1997).

in the Procrustes distortion as the distinctiveness continues to grow.

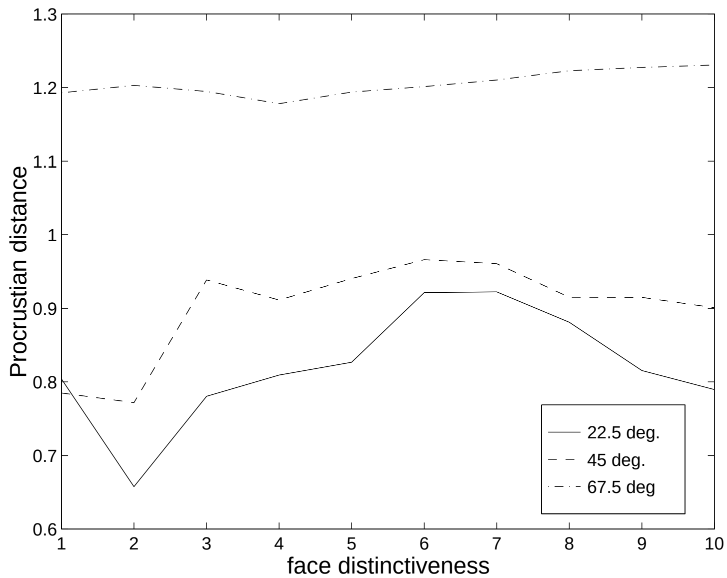


Figure 12: Procrustes distance between original and view-mapped faces as a function of face distinctiveness version and view-map condition.

Distinctiveness and the View-space

The effects of face distinctiveness on the face space representation were examined by projecting novel faces onto a set of reference faces and analyzing the resulting representations. We used 40 faces to train a Radial Basis Function (RBF) network. These reference faces were interleaved by distinctiveness (i.e., every other face; 1, 3, 5, ..., 11, 13, etc.), comprising 5 out of the 10 distinctiveness versions of the 8 original faces. The remaining 40 faces served for testing and were projected into the face space spanned by the responses of the reference-face RBF modules.

To assess the effects of face distinctiveness on the discriminability of novel faces projected into the face space, we plotted the corresponding projections directly, for different levels of distinctiveness (Figure 13). As expected, the face projections show maxima along the diagonals, due to the fact that these novel test faces were "neighbors" in the distinctiveness space to the learned faces. The extent to which there is activation off the diagonals is an indication that the model projections are confusable with other "nontarget" faces. The plotted data can be seen, therefore, to represent a confusion table of sorts. Note, first, that the relatively higher activation levels on the diagonal indicate that the similarity of the test faces to their neighbors in the learned set was sufficient to activate the RBF nodes of the learned neighbors. Of more direct interest, however, is the decrease in off-diagonal activation in the projection patterns for our parametrically more distinct face versions, effectively indicating lesser confusability of the distinct faces with other faces.

Discussion

We presented two models of generalization in face recognition, whose common main operating principle is class-based processing. The models rely on the tight similarity structure of faces considered as a class of shapes. This, in turn, gives rise to a highly structured pattern of changes precipitated by viewing transformations in a generic face representation space, as illustrated schematically in Figure 4. We would expect both of these models to work for the class-based problems considered recently by Gauthier and Tarr (1997) and Tarr and Gauthier (1998). Both of the present models learn the pattern of changes from examples, and, as a result, acquire the capability to process novel instances of a familiar class in an intelligent manner. The first model, described in section **??**, was intended as a proof of the computational feasibility of class-based processing by the simplest possible means (within the multiple-view approach to representation). The second, more elaborate model

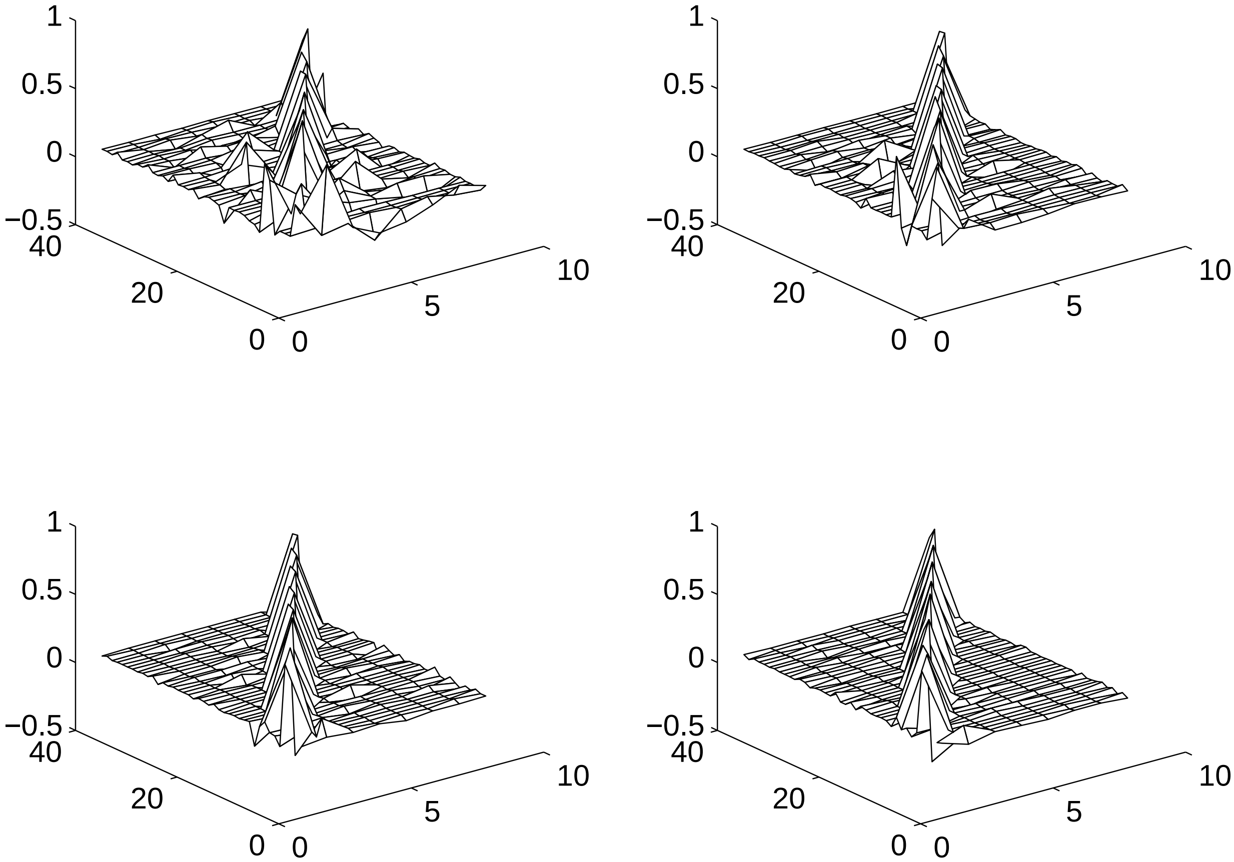


Figure 13: Face space projections for four levels of face distinctiveness, top left least distinctive, top right second most distinctive, etc. (the plot for the fifth level of distinctiveness, omitted to save space, was similar to the fourth one).

aimed at examining not only the fundamentals of class-based generalization, but also the effects of a statistical characterization of face stimuli along the typicality – distinctiveness dimension, long studied by psychologists in conjunction with generalization.

It is worth noting that because the primary components of these two models

In computer vision, at least two different computational approaches to class-based processing have been suggested recently. The first of these (Basri, 1992) concentrates on the relationship between classification and recognition, and assumes the availability of a library of 3D models of prototypical objects. The second approach (Poggio & Vetter, 1992) relies on learning class-specific transformations of linear classes of objects (linear combinations of "clouds" of points in 3D) from examples of 2D views. Because of linearity, a transformed member of a linear class is a weighted sum of similarly transformed basis objects with the same coefficients, and the same relationship holds for object views. In related work, Beymer et al. (Beymer, Shashua & Poggio, 1993) proposed to determine the transformation that relates two images of a face using an optic flow algorithm, and to apply this transformation to generate a similarly transformed image of novel face, from a single available view. In comparison with these approaches, our models rely to a great extent on the statistical characteristics of the class of faces, as, presumably, do human observers.

To substantiate the comparison between the performance of our models and that of the human observers, we conducted a series of simulations involving a controlled set of face-like stimuli. The results of these simulations, as well as reports of related studies in the psychology of face processing, offer some insights into the importance of experience in controlling the efficiency with which individual faces can be recognized by multiple-view computational models across changes in viewpoint. Particularly relevant is a recent study of the trade-off between competing constraints of recognition and viewpoint generalization for faces (O'Toole, Edelman & Bülthoff, 1998). In that study, measurements of the accuracy with which human subjects recognized individual faces across changes in viewpoint, and measures of the perceived typicality of faces were assessed. Using factor analysis, the face recognizability measures were combined with two face measures derived from the computational model: a.) the view-map quality (i.e., the quality of the estimate of the normalized face view); and, b.) the confusability of the face with other faces in the view-space. The primary finding was that human and model recognizability strongly inter-related – faces well-recognized by the model were well-recognized by human subjects. Additionally, the human typicality judgments related to the quality of the model view-map, such that faces judged to be typical by human subjects were view-mapped by the model with the higher accuracy than faces judged to be atypical. Finally, the model and human recognizability measures related inversely to the human typicality ratings and model view-map quality. Specifically, both the human subjects and the model recognized atypical faces (with poor view-map quality) more accurately than typical faces (with good view-map quality).

In summary, computational experiments outlined above, as well as psychological data found in the literature, indicate that models involving multiple-view representations of objects can learn to generalize — that is, make sense of novel views, as well as novel shapes — insofar as they can utilize class-based information. In other words, better generalization is obtained with objects that closely resemble other objects, as, indeed, is the case in face recognition by human subjects.

Notes and Acknowledgments

Some of the findings mentioned in this review resulted from S.E.'s collaboration with Maria Lando, with Yael Moses and Shimon Ullman, and with Sharon Duvdevani-Bar. AO gratefully acknowledges support from NIMH grant 1R29MH5176501A4. CORRESPONDENCE. Addressed to Shimon Edelman School of Cognitive and Computing Sciences, University of Sussex at Brighton, Falmer, BN1 9QH, UK shimone@cogs.susx.ac.uk.

References

Alfeld, P. (1989). Scattered data interpolation in three or more variables. In T. Lyche & L. Schumaker (Eds.), *Mathematical Methods in Computer Aided Geometric Design* (pp. 1–33). New York: Academic Press.

Atick, J. J., Griffin, P. A. & Redlich, A. N. (1996). The vocabulary of shape: principal shapes for probing perception and neural response. *Network*, *7*, 1–5.

Bahrick, H. P., Bahrick, P. O. & Wittlinger, R. P. (1975). Fifty years of memory for names and faces : A cross-sectional approach. *Journal of Experimental Psychology : General*, *104*, 54–75.

Bartlett, J. C. & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*, 281–290.

Basri, R. (1992). Recognition by prototypes. A.I. Memo No. 1391, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Beymer, D., Shashua, A. & Poggio, T. (1993). Example based image analysis and synthesis. A.I. Memo No. 1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Borg, I. & Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis*. Berlin: Springer.

Brigham, J. C. (1986). The influence of race on face recognition. In H. D. Ellis, M. A.

Jeeves & F. Newcombe (Eds.), *Aspects of face processing* (pp. 170–177). Dordrecht: Martinus Nijhoff.

Bülthoff, H. H. & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, *89*, 60–64.

Carey, S., Diamond, R. & Woods, B. (1980). Development of face recognition-a maturational component? *Developmental Psychology*, *16*, 257–269.

Cavanagh, P. (1995). Vision is getting easier every day. *Perception*, *24*, 1227–1232. Guest editorial.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of experimental psychology*, *115(2)*, 107–117.

Edelman, S. (1995a). Representation of similarity in 3D object discrimination. *Neural Computation*, *7*, 407–422.

Edelman, S. (1995b). Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines*, *5*, 45–68.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Science*, *1*, 296–304.

Edelman, S. & Duvdevani-Bar, S. (1997a). Similarity-based viewspace interpolation and the categorization of 3D objects. In *Proc. Similarity and Categorization Workshop* (pp. 75–81). Dept. of AI, University of Edinburgh.

Edelman, S. & Duvdevani-Bar, S. (1997b). Similarity, connectionism, and the problem of representation in vision. *Neural Computation*, *9*, 701–720.

Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.

Galper, R. E. & Hochberg, J. (1971). Recognition memory for photographs of faces. *American Journal of Psychology*, *84*, 351–354.

Gauthier, I. & Tarr, M. J. (1997). Becoming a 'Greeble' expert: Exploring the face recognition mechanism. *Vision Research*, *37*, 1673–1682.

Gordon, W. J. & Wixom, J. A. (1978). Shepard's method of 'Metric Interpolation' to bivariate and multivariate interpolation. *Mathematics of Computation*, *32*, 253–264.

Hill, H. & Bruce, V. (1994). Independent effects of lighting, orientation and stereopsis on the hollow face illusion. *Perception*, *22*, 887–897.

Hill, H. & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 986–1004.

Howell, A. J. & Buxton, H. (1995). Receptive field functions for face recognition. In *Proc. 2nd Int. Workshop on Parallel Modelling of Neural Operators for Pattern Recognition (PAMONOP)* (pp. 83–92). Faro, Portugal.

Lando, M. & Edelman, S. (1995). Receptive field spaces and class-based generalization from a single view in face recognition. *Network*, *6*, 551–576.

Light, L. L., Kayra-Stuart, F. & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 212–228.

Logothetis, N. & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, *3*, 270–288.

Logothetis, N. K., Pauls, J. & Poggio, T. (1995). Shape recognition in the inferior temporal cortex of monkeys. *Current Biology*, *5*, 552–563.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

Moses, Y., Adini, Y., & Ullman, S. (1994). Face recognition: the problem of compensating for illumination changes. In Eklundh, J.-O. (Ed.), *Proc. ECCV-94* (pp. 286–296). Springer-Verlag.

Moses, Y., Ullman, S. & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, *25*, 443–462.

Newell, F., Chiroro, P. & Valentine, T. (1996). Recognising unfamiliar faces: The effects of distinctiveness and view. *Q. J. Exp. Psychol.* submitted.

O'Toole, A., Deffenbacher, K., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory and Cognition*, *22*, 208–224.

O'Toole, A., Vetter, T., Troje, N. & Bülthoff, H. H. (1997). Sex classification is better with three-dimensional head structure than with image intensity information. *Perception*, *26*, 75–84.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A. & Valentin, D. (1993). Low dimensional representation of faces in high dimensions of the space. *Journal of the Optical Society of America*, *10*, 405–410.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A. & Valentin, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine (Ed.), *Cognitive and Computational Aspects of Face Recognition* (pp. 159–182). New York, NY: Routledge.

O'Toole, A. J. & Edelman, S. (1996). Face distinctiveness in recognition across viewpoint: An analysis of the statistical structure of face spaces. In Essa, I. (Ed.), *Proc. 2nd Intl. Conf. on Face and Gesture Recognition* (pp. 10–15).

O'Toole, A. J., Edelman, S. & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, –. in press.

Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.

Poggio, T. & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, *247*, 978–982.

Poggio, T. & Vetter, T. (1992). Recognition and structure from one 2D model view: observations on prototypes, object classes, and symmetries. A.I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Rhodes, G. & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: towards a model of configural coding. *Perception*, *19*, 773–794.

Rock, I. (1973). *Orientation and form.* Cambridge, MA: MIT Press.

Rock, I. (1988). On thompson's inverted face phenomenon. *Perception*, *17*, 815–817.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly spaced data. In *Proc. 23rd National Conference ACM* (pp. 517–524). ACM.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, *9*, 483–484.

Townsend, J. T. & Thomas, R. (1993). On the need for a general quantitative theory of pattern similarity. In S. C. Masin (Ed.), *Foundations of perceptual theory¿ Advances in psychology: Vol 99* (pp. 297–368). Amsterdam: North-Holland/Elsevier Science Publishers.

Ullman, S. (1996). *High level vision.* Cambridge, MA: MIT Press.

Valentine, T. (1991). Representation and process in face recognition. In R. Watt (Ed.), *Vision and visual dysfunction*, Volume 14 chapter 9, (pp. 107–124). London: Macmillan.

Valentine, T. & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*, 525–535.

Weiss, Y. & Edelman, S. (1995). Representation of similarity as a goal of early visual processing. *Network*, *6*, 19–41.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.