

---

## Chapter X. Psychological and Neural Perspectives on Human Face Recognition

Alice J. O'Toole<sup>1</sup>

School of Behavioral and Brain Sciences, University of Texas at Dallas

Human face “processing” skills can make simultaneous use of a variety of information in the face, including information about the age, sex, race, identity, and even current mood of the person. We are further able to track facial motions that alter the configuration of features, making it difficult to encode the structure of the face. Facial movements include the gestures we make when we speak, changes in our gaze direction or head pose, and expressions like smiling and frowning. These movements play an important role in human social interactions and in the survival mechanisms that protect us from danger, indicate the presence of a potential mate, and direct our attention to an object or event of importance in our environment.

We are capable of these impressive feats of visual information processing even when viewing conditions are variable or less than optimal — for example, when the face is poorly illuminated, viewed from an unusual vantage point or viewed from a distance so that the resolution of the image on the retina is limited. These viewing parameters have proven challenging for computational models of face recognition.

Given the evolutionary importance of accurate and speedy recognition of human faces and the recognition of the social information conveyed by expressions and movements, it is perhaps not surprising that the neural processing of faces has been studied intensively in recent decades. What is known to date, indicates that several areas of the human brain are involved in the analysis of the human face and that these areas may distinguish processing according to the functions of information they analyze. We will see that the analysis of the static features of faces, which convey identity and categorical information about faces, is probably carried out in a different part of the brain than the analysis of the motions that carry social information. The processing of emotional information from the face is further differentiated neurally.

From a combined psychophysical and neural perspective, human face recognition serves as an example to the developers of automatic face recognition algorithms that it is possible and indeed “easy” to recognize faces, even when viewing conditions are challenging. The human system, however, is not infallible. Errors of identification abound in circumstances that challenge the human system on its own terms. The purpose of this chapter is to provide an overview of the human face processing sys-

tem from both a psychological and neural perspective. We hope that this overview and analysis of the human system may provide insights into successful strategies for dealing with the problem of automatic face recognition. We further hope that it will provide a useful guide for comparing the strengths and weaknesses of the human system with those of currently available face recognition and analysis algorithms.

In this chapter, we consider the psychological and neural aspects of face perception and recognition. For the psychological part, we first discuss the diversity of tasks humans perform with faces and link these to the kinds of information that supports each task. Next, we consider the characteristics of human face recognition. In this section we address questions about the circumstances under which humans excel at face recognition and the circumstances under which recognition accuracy begins to fail. More specifically, we present an overview of that factors that affect human memory for faces. For the neural part of the chapter, we consider what is known about the neural processing of faces. We present a multiple systems model of neural processing that suggests a functional organization of facial analysis. This analysis distinguishes among the different components of facial features and motions that subserve various tasks. Finally, we will summarize the points of interest for algorithm developers in seeking solutions to the challenges of robust and accurate face recognition.

## 1 Psychological Aspects of Face Perception and Recognition

### 1.1 Extracting information from the human face

Perhaps the most remarkable aspect of the human face is the diversity of information it provides to the human observer, both perceptually and socially. We consider, in turn, the major tasks that can be accomplished with this information.

#### Identity

Each human face is unique, and as such, provides information about the identity of its owner. Humans can keep track of hundreds (if not thousands) of individual faces. This far exceeds our ability to memorize individual exemplars from any other class of objects (e.g., How many individual suitcases can we remember?).

To identify a face, we must locate and encode the information that makes the face *unique* or different from all other faces we have seen before and from all other unknown faces. As impressive as it is to be able to identify a face we have seen before, it is equally impressive to state with confidence that a face is one we have never seen before.

Computational models like principal component analysis have given insight into the nature of the information in faces that makes them unique. In a series of simulations, face recognition accuracy for sets of eigenvectors was found to be *highest* for eigenvectors with relatively *low* eigenvalues [1]. This may seem surprising from an engineering point of view, which suggests that low dimensional representations

should be based on the eigenvectors with the largest eigenvalues. From a perceptual point of view, however, the finding makes sense. Eigenvectors with relatively small eigenvalues explain little variance in the set of faces. Indeed, to identify a face, we need to encode the information that a face shares with few other faces in the set. This information will be captured in eigenvectors with small eigenvalues. A perceptual illustration of the finding occurs in Figure 1. The original face appears on the left. The middle image is a reconstruction of the face using the first 40 eigenvectors. The rightmost image is reconstruction of the face with all but the first 40 eigenvectors. As can be seen, it is much easier to identify the face when it is reconstructed with eigenvectors with relatively smaller eigenvalues.



**Fig. 1.** An illustration of the identity-specific information in faces, using principal component analysis [1]. The original face appears on the left. The middle image is a reconstruction of the face using the first 40 eigenvectors. The rightmost image is reconstruction of the face with all but the first 40 eigenvectors. As can be seen, it is much easier to identify the face when it is reconstructed with eigenvectors with relatively smaller eigenvalues.

Because faces all share the same set of “features” (eyes, nose, mouth, etc.) arranged in roughly the same configuration, the information that makes individual faces unique must be found in subtle variations in the form and configuration of the facial features. Data from human memory experiments suggest that humans use both feature-based and configural information to recognize faces (e.g., [2]), with perhaps special reliance on facial configurations (e.g., [3]). The reliance of the human perceptual system on configural information has been demonstrated using various experimental manipulations. These are aimed at perturbing the configuration of a face or at disrupting our ability to process the configural information in a face. The manipulations used in previous work include distortions of the relative positions of the mouth, eyes, and nose [2], inverting a face [4], and altering the vertical alignment of the contours [5]. All of these manipulations strongly impact human recognition accuracy and processing speed.

An excellent example of the importance of configuration in human face perception can be illustrated with the classic “Thatcher illusion”, so named because it was demonstrated first with Margaret Thatcher’s face [6]. The illusion is illustrated in Figure 2. A face can be “Thatcherized” by inverting the eyes and the mouth,

and then inverting the entire picture. Most people do not notice anything peculiar about the inverted face. Upright, however, we see a gross distortion of the configuration of the facial features. There is evidence that humans are highly sensitive to the configuration of the features in a face, but that the processing of configuration is limited to faces presented in an upright orientation. The phenomenon illustrates that human face perception has some important processing limits for nontypical views.



**Fig. 2.** Illustration of the Thatcher Illusion. The inverted face appears normal. Upright, however, the configural distortion is evident. The illusion illustrates the limits of configural processing for atypical views of the face.

In the human memory literature, the terms *identification* and *recognition* are often distinguished. Recognition is the process by which a human observer judges whether or not a face has been seen before. This definition includes the two components of accuracy required for face recognition. First, a previously encountered face is “recognized”, when it is judged as familiar or “known”. Second, a novel or previously unseen face is “correctly rejected”, when it is judged as unfamiliar or “unknown”. Correct judgments on the former component are referred to as “hits” and mistakes are referred to as “misses”. Correct judgments on the latter component are referred to as “correct rejections” and mistakes are referred to as “false alarms”. Recognition memory performance is generally measured using the signal detection theory measure of  $d'$ , which considers both hits and false alarms [7].

An important characteristic of human face recognition is that it can occur in the absence of the ability to *identify* a face. For example, one may be absolutely certain that they recognize the grocery store clerk but may fail to retrieve a name or context of encounter (e.g., “He is in one of my classes” or “That’s Bob from the gym”). The feeling of certainty that we recognize a face is, therefore, not linked inextricably to the memory of an individual person. This is a common perceptual phenomenon that occurs with stimuli in other modalities (e.g., a familiar tune or a familiar taste that is recognized but not identified).

In the human memory literature, identification presumes recognition but requires the retrieval of a semantic label like a name or context of encounter. For face recognition algorithms, identification is the task most commonly performed. The

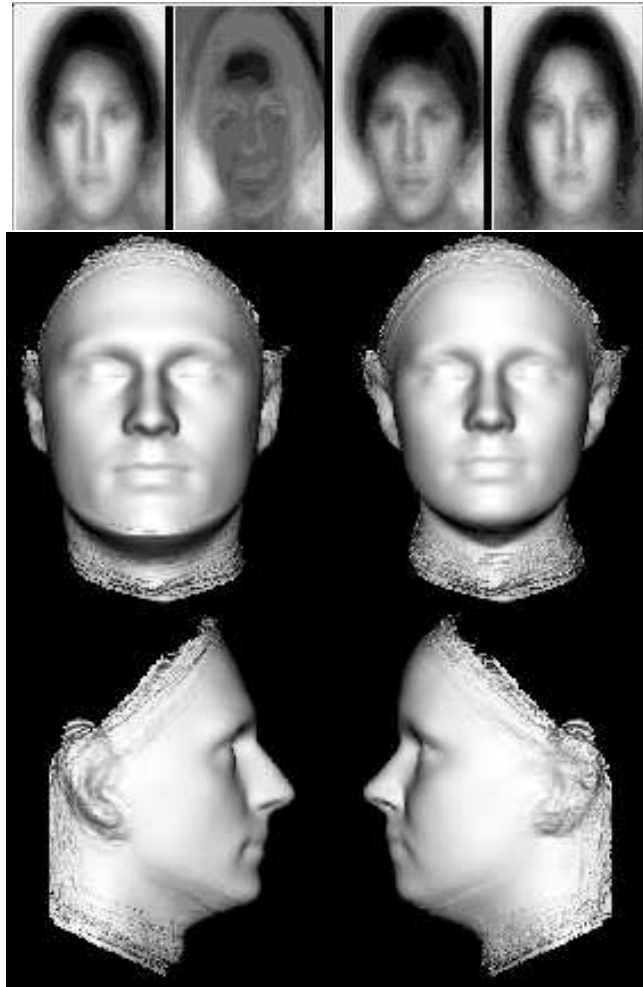
retrieval process used by most algorithms is specific to an individual face rather than being a general judgment about whether or not the face is “in the database somewhere”. Only a few algorithms allow for recognition to occur independently of identification, and these tend to be models aimed at simulating human performance (e.g. [8]).

### Visually-derived Semantic Categories of Faces

In addition to our ability to recognize and identify faces, humans can also categorize faces along a number of dimensions referred to as “visually derived semantic categories” [10]. These categories include race, sex, and age. By a broader definition, one can also include other visually specified, albeit abstract, categories like personality characteristics. For example, most humans will happily make a judgment about whether or not a face looks “generous” or “extroverted”. Faces can be categorized quickly and easily on all of these dimensions. An intriguing aspect of this phenomenon is that making such judgments actually *increases* human accuracy by comparison to making physical feature-based judgments (e.g., nose size) (see [11]).

In contrast to the information needed to specify facial identity, (i.e., what makes a face unique or different from all others), visually derived semantic categorizations are based on the features that a face shares with an entire category of faces. To determine that a face is male, for example, we must locate and encode the features that the face shares with other male faces. As was the case for identity-specific information, computational models like principal component analysis can provide insight into the nature of category-specific information in the face. Simulations looking at gender classification accuracy for individual eigenvectors showed that eigenvectors with relatively large eigenvalues contain the most useful information for this task [1]. An illustration of this appears in Figure 3. The top row (leftmost) shows the first eigenvector from a principal component analysis from which the average face was not subtracted. This image approximates the average face. Next to it appears the second eigenvector. The images on the right were constructed by adding the second eigenvector to the average and by subtracting the second eigenvector from the average, respectively. The general male-female forms are contrasted in these combined images. Using the eigenvector weights for the second eigenvector alone, achieved accuracy levels over 75 percent. An analogous demonstration using three-dimensional laser scans appears at the bottom of the figure [14]. This shows the average combined positively and negatively with the first eigenvector, which was the best predictor of the gender of the face.

There has been less research on the perception of visually derived semantic categories than on face recognition. Notwithstanding, research on the perception of face gender indicates that humans are highly accurate at sex classification [12, 13, 14], even when obvious “surface cues” like facial and head hair are absent. They are also capable of making gender judgments on the faces of young children, which contain more subtle cues to gender than adult faces [15]. Male and female prototypes of this information appear in Figure 4. These were constructed by morphing pairs of boy’s (or girl’s) faces together, and then morphing together pairs of the pairs,



**Fig. 3.** Illustration of the computationally-derived information in images [1] and three dimensional head models [14] that specifies the gender of a face. The top part of the figure shows the results of a principal component analysis of face images. Left to right, we see the average face, the second eigenvector, the average face plus the second eigenvector, and the average face minus the second eigenvector. Face projections onto this eigenvector predicted the sex of the face very accurately. The bottom part of the figure shows an analogous analysis for laser scanned heads. In the top row, the average plus and minus the first eigenvector are displayed. In the bottom row, the average plus and minus the sixth eigenvector are displayed.

etc., until reaching a single convergent image was reached [15]. Humans are also surprisingly good at estimating the age of a face ([16, 17]).



**Fig. 4.** Illustration of the computationally-derived gender information in children's faces. Left is the male prototype, made by morphing boys together (see text for details), and right is the female prototype.

### Facial Expressions, Movement, and Social Signals

The human face moves and deforms in a variety of ways when we speak or display facial expressions. We can also orient ourselves within our environment by moving our head or eyes. We interpret these expressions and movements quickly and accurately. Virtually all of the face and head movements convey a social message. Rolling our eyes as we speak adds an element of disbelief or skepticism to what we are saying. Expressions of fear, happiness, disgust, and anger are readily and universally interpretable as conveying information about the internal state of another person [18]. Finally, the head movements that accompany changes in the focus of our attention provide cues that signal the beginning and end of social interactions.

There is limited data in the psychological literature linking facial movements to social interpretations. In general, except at the extremes, facial expressions and gestures are very difficult to produce on demand and are further difficult to quantify as stimuli. There is also only limited ground truth available with facial expressions. This makes controlled experimentation difficult. Anecdotally, however, it is clear that these movements provide constant feedback to the perceiver that helps to guide and structure a social interaction.

It is possible, but not certain, that facial movements may also complicate the job of the perceiver for recognizing faces. It is likely that the information needed to recognize faces can be found in the invariant or unique form and configuration of the features. Non-rigid facial movements alter the configuration of facial features, often in a dramatic way. Research on the effects of various motions on recognition accuracy is just beginning, and complete answers to these questions are not yet available. A detailed discussion of these issues can be found in a recent review [19].

## 1.2 Characteristics of Human Face Recognition

Human face recognition accuracy varies as a function of *stimulus factors*, *subject factors*, and *photometric conditions*. All three of these factors, (including subject factors!) are relevant for predicting the accuracy of automatic face recognition algorithms. We will explain this point in more detail shortly. For humans, and perhaps also for machines, familiarity with the face can be an important predictor of the robustness of face recognition over changes in viewing parameters.

### Stimulus Factors

Not all faces are recognized equally accurately. Indeed, some people have highly unusual faces, with very distinctive features or configurations of features. These faces seem, and are, easy to remember. Specifically, distinctive faces elicit more hits and fewer false alarms than more typical faces, which have relatively few distinguishing characteristics. The negative correlation between the typicality and recognizability of faces is one of the most robust findings in the face recognition literature (e.g., [20]). The finding is relevant for predicting face recognition success for human observers at the level of individual faces.

More theoretically, the relationship between face typicality and recognizability has interesting implications for understanding the way human face recognition works. When a human observer judges a face to be unusual or distinctive, nominally it might be because “the nose is too big”, or because “the eyes are too close together”. It is clear however, that implicit in these judgments is a reference to internalized knowledge about how long a nose *should be* or how close together eyes *should be*. The typicality-recognizability finding has been interpreted as evidence that human observers store a representation of the average or *prototype* face, against which all other faces are compared [21]. This suggests that individual faces are represented, not in absolute terms, but in relative terms.

The typicality-recognizability relationship suggests that individual faces may be represented in human memory in terms of their deviation from the average. There are interesting computational models of face encoding and synthesis that share this kind of encoding. Specifically, algorithms that directly use the correspondence of features to an average ([22, 25, 26], share the basic principals of a prototype theory of human face recognition, because the faces are encoded relative to one another via the prototype or average of the faces.

At a more concrete level, the prototype theory of face recognition has been modeled in the context of a multidimensional face space [21]. By this account, individual faces can be thought of metaphorically as points or vectors in the space, with the axes of this space representing the features on which faces vary. The prototype or average face is at the center of the space. The face space model predicts the typicality-recognizability relationship by assuming that the density of faces is highest near the center of the space, and falls off as a function of the distance from the center. Typical faces, close to the center of the space, are easily confused with

other faces, yielding a high probability of false alarms. Distinctive faces are found further from the center of the space, and are not easily confused with other faces.

Face space models have been implemented computationally as well and are currently the most common base for automatic face recognition algorithms. For example, the principal components model or *eigenface model* [27] implements the the psychological face space model in a concrete fashion. Principal components form the statistically-based feature axes in the space. The coordinates of individual faces with respect to the axes locate individual faces in the space. Principal component-based face recognition models can be shown to predict aspects of the typicality-recognizability relationship for human observers [8].

The typicality-recognizability finding also suggests a reason for the superior recognition of caricatures over veridical faces [9]. Artists draw caricatures in a way that exaggerates the distinctive features in a face. For example, Mick Jagger can be caricatured by making his already thick lips, even thicker. Computer-based caricatures are made by comparing the feature values for an individual face with feature values for an average face, and then redrawing the face, exaggerating features that deviate substantially from the average (see [17, 9] for a review). Several studies have shown that computer-generated caricatures are recognized more accurately and more quickly than veridical versions of the faces. This is a notable finding in that caricatures are clearly distorted versions of faces we have experienced. Thus, the finding suggests that a distorted version of a face is actually easier to recognize than the exact template stored.

A sophisticated method for producing caricatures from laser scans was developed recently [22]. The laser scans are put into point-by-point correspondence [22] with an algorithm based on optic flow analysis. Once in correspondance, it is possible to create caricatures by simply by increasing the distance from the average face in the face space, and reconstructing the laser scan with the new “feature values”. Note that features here means complex face shapes. An illustration of this appears in Figure 5, created using Blanz and Vetter’s caricature generator for laser scan data [22]. The representation here includes both the surface map and the overlying texture map. It is clear that the distinctive and unique aspects of this individual are enhanced and exaggerated in the caricatured version of the face.



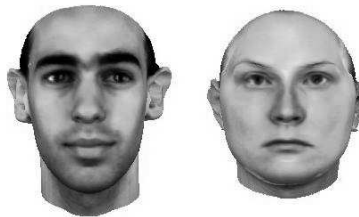
**Fig. 5.** Illustration of computer-generated three-dimensional caricatures from Blanz and Vetter’s caricature generator [22]. Both the surface and texture data are caricatured.

In fact, it is possible to create and display an entire trajectory of faces through the space, beginning at the average face and going toward the original (see Figure 6). The faces in between are called “anti-caricatures” and are recognized less accurately by humans than are veridical faces (see [17] for a review of studies).



**Fig. 6.** Illustration of computer-generated three-dimensional anti-caricatures from Blanz and Vetter’s caricature generator [22]. The average face is on the left and the original face is on the right. The faces in between lie along a trajectory from the average to original face in face space.

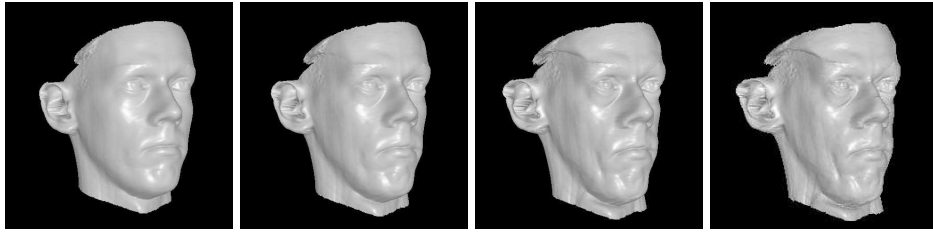
The concept of a trajectory can be extended even further by imagining what a face would look like “on the other side of the mean”. Thus, if we extend the trajectory in Figure 5 in the *opposite direction*, i.e., through the mean and out the other side, we would arrive at the “opposite” of the face (see Figure 7). This “anti-face” is one in which all of the “feature values” are inverted. So, the dark-skinned, dark-eyed, thin face becomes a light-skinned, light-eyed, round face.[23]. The anti-face is linked perceptually to the face in an opponent fashion, suggesting that faces may be stored in the brain in relation to the existence of a prototype [24]. Staring at the anti-face for a short amount of time seems to facilitate the identification of the real face [24]. This situation suggests a kind of excitory-inhibitory trade-off between feature values around the average.



**Fig. 7.** Illustration of computer-generated three-dimensional anti-caricatures from Blanz and Vetter’s caricature generator [22]. The veridical face is on the left and the *anti-face* is on the right. The anti-face lies on the other side of the mean, along a trajectory from the original face, through the mean, and out the other side [23].

There are also interesting additional perceptual dimensions to the caricatured faces. When a computer-based algorithm is applied to the three-dimensional shape

information from laser scans omitting the surface texture, caricaturing actually increases the perceived *age* of the face [17]. An illustration appears in Figure 8. Again, the laser scans are put into correspondence [22] and are exaggerated simply by increasing the distance from the average face in the face space. Human subjects recognize the caricatured faces more accurately and judge these caricatured faces to be older (even decades older!) than the original veridical face [17].



**Fig. 8.** Illustration of computer-generated three-dimensional caricatures from Blanz and Vetter’s caricature generator [22]. When caricaturing is applied to three-dimensional head models, the face appears to age.

The robust relationship between the typicality of a face and its recognizability has practical implications for the accuracy of eyewitness identifications. Specifically, if we consider the effects of typicality, face recognition accuracy should be expected to vary with the face itself. In an eyewitness identification situation, therefore, typical suspects are more likely to be identified erroneously than distinctive suspects. This error bias for typical faces is also likely to be the case for automatic face recognition algorithms, many of which are based on statistical or multidimensional space based models of faces [8].

### Interaction of Stimulus and Subject Factors

The interaction of stimulus and subject factors in face recognition should be expected when one takes into account the likelihood that we encode faces in terms of their deviation from the average or prototype face. It would seem highly unlikely that the “average” face, which we learn and internalize as we begin to experience faces, is the same for all groups of subjects. For example, the faces we might encounter growing up in Japan would yield a very different average than those we might encounter growing up in India, or the United States. In fact, as is well known anecdotally, there is good psychophysical evidence for an *other-race effect* [28, 29, 8]. The other-race effect is the phenomenon that we recognize faces of our own race more accurately than faces of other races.

There are a number of explanations for the other race effect, but all involve the idea that there is an interaction between stimulus factors and subject experience factors. The *contact hypothesis* suggests that a subject’s experience with faces of their own race biases for the encoding of features that are most useful for distinguishing among own-race race. This enables subjects to create a detailed and

accurate representation of the distinctive features of own-race faces. This causes a problem with other-race faces however, because they are not well-characterized by these features. A simple-minded example might go as follows. Caucasians might rely on eye color as an important facial feature. Although this would be a helpful feature for Caucasian faces, it is likely to be far less helpful for Asian faces. A perceptual consequence of the failure to accurately encode the distinguishing facial characteristics of other-race faces is that these faces will be perceived as more similar, one to the next, than own-race faces. This yields the well-known feeling, and oft-quoted statement, that other-race faces “all look alike to me”.

The other-race effect for humans is surprisingly relevant for computational algorithms of face recognition, which often rely on statistically-based learning procedures. These procedures are used commonly for acquiring the basis set or principal components with which faces are encoded. It is clear that the racial composition of the training set of faces will impact the performance of these algorithms depending on the race of the target face.

In a recent study, 13 automatic face recognition algorithms were tested for the presence of an other-race effect [30]. The results were interesting and in some ways surprising. First, algorithms based on generic PCA actually performed better on faces in the “minority race” than in the “majority race”. Minority and majority refer to the relative numbers of faces of two races in the training set. This is because minority race faces are distinctive or unusual relative to the other faces in the database. These distinctive faces inhabit a less populated area of the face space and are thus less likely to have competing neighbors which might be mistaken for them. A second set of algorithms that linked PCA to a second stage learning algorithm (such as Fischer discriminant analysis) showed the classic other-race effect, with performance better for majority race faces. This is likely to be due to the fact that the second stage training procedures serve to warp the space to maximize the distance between different faces in the space. This improves majority race face encoding at the cost of minority race face encoding.

### **Face Representations, Photometric Factors and Familiarity**

Much recent research has been devoted to understanding how humans are able to recognize objects and faces when there are changes in the photometric conditions between learning and test stimuli. The rationale for this research is as follows. Studying how well we generalize recognition of faces to novel viewing conditions may give insight into the kinds of representations our brains create of faces. This can occur via inferences about the way learned information can be used to recognize altered versions of a target stimulus. Thus, some representations (e.g., an object-centered three-dimensional representation) predict good generalization to novel views and illumination conditions. For this reason, the effects of photometric variables, like changes in the illumination or viewpoint, have been studied in some detail. We will consider only the effects of viewpoint and illumination change, because they currently represent one of the most important challenges for algorithms

to be able to function in real world conditions. In any case, the effects for other photometric changes are quite consistent with illumination and viewpoint effects.

We first present a brief overview of the psychological theories of face and object representation. We then discuss the relevant data and its implications for the nature of face representations.

### Face Representation Debate

There are long-standing questions in the psychological literature about the nature of face representations. Much of this debate concerns differences in the extent to which psychologists posit two- versus three-dimensional representations of faces and objects. It is worth noting that these theories have been developed primarily to account for object recognition. There are important differences between face and object recognition, including the level of analysis required for tasks. Object recognition refers to the ability to classify individual objects as members of a category (e.g., “That is a chair”). Face recognition refers usually to a recognition or identity decision about a single exemplar of the face category (e.g., “I’ve met you before”, “There’s Bob!”, respectively). Recent developments on refining computational models to simultaneously subserve performance at both of these levels can be found in a recent review [31]. For present purposes, we will discuss only the history and rationale of the two competing approaches.

*Structural theories* suggest that recognition occurs by building a three dimensional representation of faces and objects from the two dimensional images that appear on the retina. The most well-known of these theories is the “recognition by components” theory [32], based on the foundations laid in Marr’s classic book on vision [33].

*Interpolation-based models* like those advanced by Poggio and Edelman [34] posit two-dimensional representations of objects. By this theory, we encode multiple view-based representations of faces and objects. Recognition of a novel view of an object or face occurs by interpolation to the closest previously seen view. Compromise or hybrid accounts of recognition posit a correspondence process by which novel views are aligned to view-based templates and recognition occurs by a matching process [35]. More recent models have tackled the complex problem of affine transformations with biologically plausible computational schemes, but these models retain the basic view-based nature of the encoding [31].

### Familiarity and Face Recognition over Photometric Inconsistencies

Under some circumstances, humans show a remarkable capacity to recognize people under very poor viewing conditions. We can all recognize the face of a friend from a single glance on a dark train platform or in a blurry low quality photograph. The psychological literature is clear, however, that this ability is limited to faces with which we have previous experience or familiarity. (For a recent review of the literature on familiar face recognition, see [36]).

When psychologists discuss recognition of *unfamiliar faces*, they usually mean recognition of a face with which we have had only one previous encounter. This is the kind of performance we might expect from a security guard who is interested in recognizing people from the memory of a photograph he has viewed previously. Most experiments in the laboratory are done on relatively unfamiliar faces, due to the difficulties encountered in testing memory for familiar faces. For example, a familiar face recognition experiment requires a set of stimulus faces known to all subjects in the experiment (such as famous faces) or requires the construction of stimulus sets tailored to the experience of each individual subject. These difficulties in experimentation are the reason that the data on familiar face recognition are limited.

For unfamiliar faces, there is general agreement that recognition declines as a function of the difference between learning and testing conditions. (For a recent review of unfamiliar face recognition, see [37]). In most controlled human memory experiments, subjects learn previously unfamiliar faces from a single image and are asked to recognize the face from a second novel image of the person after a delay. When there are changes in the viewpoint between learning and test images, such as from frontal to profile views, accuracy declines as a function of the difference between learning and test images [38, 39, 40].

The difficulties over viewpoint change are seen also with perceptual matching experiments that do not involve a memory component. In matching experiments, subjects are asked to determine if the persons pictured in two simultaneously or sequentially presented images are the same or different. Match accuracy declines as a function of the difference in viewpoint [40, 41].

There are comparable deficits in recognition accuracy when the illumination conditions are changed. These too depend on the degree of change between the learning and test image. At one extreme, photographic negatives of people, which approximate the unusual condition of lighting a face from below, are very difficult to recognize [4]. But, even less extreme changes in illumination produce detrimental effects on accuracy (e.g., [42, 43]).

The debate about whether these results support a two-dimensional or three-dimensional representation of faces has evolved over the past ten years. It is reasonable to assume that *view dependence* in recognition performance supports a two-dimensional representation hypothesis, under the assumption that one matches the target with the image previously learned. The greater the difference between the two images, the worse the performance. Alternatively, however, it is possible that view dependent recognition performance may be indicative of a three-dimensional representation that is only poorly constructed due to a lack of data. In other words, a good three dimensional representation may require exposure to more than a single view to be accurate enough to support view independent performance.

In summary, there is a problem with using view independence to argue for a three dimensional hypothesis and view dependence to argue for a two dimensional hypothesis. Specifically, without familiarity, performance is usually view dependent and with experience or familiarity it is usually view independent. A two-dimensional representation is consistent with view independence provided that one has expe-

rience with multiple views of a face. Likewise, a three dimensional representation can yield view dependent performance when one has not experienced a sufficient number of views to build a good structural representation. It may, therefore, be impossible to settle this issue using only psychological data. The addition of computational and neural data, to make very precise predictions about generalization and familiarity, may be necessary to tease apart these hypotheses.

Before leaving the issue of recognition for familiar and unfamiliar faces, we note two recent experiments that speak to the robustness of familiar face recognition in a naturalistic setting. Burton and colleagues asked participants to pick out faces from poor quality videos similar to those used in low-cost security systems [44]. They used video footage of professors captured from university surveillance cameras. Burton et al. tested three groups of participants: students familiar with the professors; students unfamiliar with the professors; and a group of trained police officers who were unfamiliar with the professors. The performance of all but the familiar subjects was very poor. To track down the source of the good performance by the familiar subjects, Burton, et al. designed a second experiment to determine which aspect(s) of the stimuli contributed to the familiarity advantage. They did this by editing the videos to obscure either the body/gait or the face. Burton et al. found that the “face-obscured” version of the tape resulted in much worse recognition performance than the “body-obscured” version. The authors concluded that face information plays the key role in identifying someone familiar, even when other informative and cues like body and gait are present.

In summary, human memory research indicates that memory for familiar faces, i.e., those with which we have experience, is robust against a variety of changes in photometric conditions. For relatively unfamiliar faces, recognition performance for humans suffers as a function of the difference between learning and test images. This makes human recognition performance for unfamiliar faces similar to the performance of most face recognition algorithms, which are similarly challenged by changes in photometric factors between learning and test. It is important to bear in mind that most algorithms are limited in terms of the number and type of views and illumination conditions available to train the model (see [31] for a discussion of this issue). When human observers are similarly limited, recognition accuracy is also poor. Understanding the process and representational advantages that humans acquire as they become familiar with a face may therefore be useful for extending the capabilities of algorithms in more naturalistic viewing conditions.

## 2 Neural Systems Underlying Face Recognition

The neural systems underlying face recognition have been studied over the past 30 years from the perspective of neuropsychology, neurophysiology, and functional neuroimaging. It is beyond the scope of this chapter to provide a comprehensive review of this extensive literature. Rather, we will provide a brief sketch of the kinds of literature that have formed the foundations of the inquiry and will then

focus on a recent model that has begun to make sense of the diverse and plentiful findings from these various disciplines.

## 2.1 Diversity of Neural Data on Face Process

### Neuropsychology

Studies of patients who suffer brain injuries after accidents or stroke have often provided important insights into the organization of neural function. For faces, neuropsychological data on *prosopagnosia* provided some of the first hints that brain processing of faces might be “special” in some way. Prosopagnosia is a rare condition in which a patient, after a brain injury or stroke, loses the ability to recognize faces, despite a preserved ability to recognize other visual objects. A person with prosopagnosia can recognize his car, his house, his clothes, but fails to recognize the people he knows by their faces. Even more intriguing, some prosopagnosics can recognize facial expressions, while failing entirely to recognize the identity of the person displaying the facial expression [45].

It is worth noting that the problem in prosopagnosia is not a problem with identifying the person. Rather, prosopagnosics can identify people accurately using their voices or other cues like the kinds of clothes they wear. They simply lack the ability to encode the identity-specific information in individual faces. In other words, they fail to encode the information that makes an individual face unique.

The existence of prosopagnosia suggests some localization of function in the brain for identifying faces. This general claim has been supported in functional neuroimaging studies as well, which we will discuss shortly. It further suggests some modularity of processing. Modular processing is used to refer to tasks that are encapsulated in the brain and are thought to function as relatively independent systems. Though there is still much controversy about the extent to which the processing of faces in the brain is modular (see [46, 47]), there is general agreement that at least some of the important processes are relatively local and primarily (if not exclusively) dedicated to analyzing faces.

### Neurophysiology

Over the last three decades or so, the function of individual neurons in visual cortex of animals has been probed with single and multiple electrode recordings. Electrodes can record the activity of single neurons. Used in conjunction with a stimulus that activates the neuron, these methods can provide a catalog of the stimulus features that are encoded by neurons in the visual system. In these studies, neurophysiologists insert an electrode into an individual neuron in the brain, while stimulating the animal with a visual stimulus, e.g., a moving bar of light. The “effective visual stimulus” will cause the neurons to discharge and will define the receptive field properties of the neuron. With these methods, neurons selective for oriented lines, wavelengths, motion direction/speed, and numerous other features have been discovered in the occipital lobe of the brain. Beyond the primary visual

areas in the occipital lobe of the brain, higher level visual areas in the temporal cortex have been found that are selective for complex visual forms and objects [48, 49], including faces and hands. Some of the *face-selective neurons* respond only to particular views of faces [50].

The selectivity of individual neurons for faces has lent support to the idea that face processing in the brain may be confined to a (or some) relatively local area(s) of the brain. It further supports the idea that face analysis may act as a special purpose system in the brain. The claim that face processing is “special” in this way, however, has been controversial for a number of reasons. To begin with, unlike object recognition, face recognition requires an ability to keep track of many individual exemplars from the category of faces. In other words, we require a level of visual expertise with faces that is not needed with most (or possibly any) other category of objects.

This second definition of “special” implies the need to process faces at a level of visual sophistication beyond that required for other objects. It has been hypothesized, therefore, that regions of the brain that appear to be selective only for faces may actually be selective for the *processes* needed to achieve the high level of expertise we show for faces [51]. This hypothesis has been examined with functional neuroimaging methods by looking at the responses of face selective areas to other objects (e.g., birds) with which some subjects have perceptual expertise (e.g., bird-watchers). Indeed, there is some evidence that a particular brain region that is selective for faces, responds also to object categories with which a particular subject has perceptual expertise [51]. For reasons having to do with the complexity and non-uniformity of fMRI analyses, the question remains a topic of active debate.

Finally, we note that one must be cautious in interpreting neurophysiological data, which is always gathered using animals, usually primates, as subjects. Although the brains of primates are similar to humans, there are still important differences in the structure and function of the various brain areas. For this reason, functional neuroimaging analyses, applied to human brains, can be helpful in ascertaining critical differences between the brains of human and non-human primates.

### Functional Neuroimaging

A recent set of tools in the arsenal of brain study allows a glimpse of the human brain while it is working. Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are two of the most common functional neuroimaging tools. Although they work in very different ways, both allow a high resolution spatial brain image to be overlaid with a high resolution temporal image of the activity levels of different parts of the brain as the subject engages in a task (like viewing a face, or reading, or listening to music). Using this technology, neuroscientists have recently named a small area in the human inferior temporal lobe of the brain the “fusiform face area” or FFA [52]. This area of the brain responds maximally and selectively to the passive viewing of faces [52].

Interestingly, other areas of the brain seem to respond to faces as well, but are constrained by additional parameters of the stimulus such as expression or

facial movement. We consider these data in the context of a recent model of neural processing of faces.

## 2.2 Multiple Systems Model

As noted, the brain systems that process information about human faces have been studied for many decades using single unit neurophysiology in primates and neuropsychological case studies of prosopagnosia. Neuroimaging studies from the past decade have enriched and extended our knowledge of the complex neural processing that underlies face perception. These studies allow researchers to examine the normal human brain as it carries out face processing tasks. The number of brain areas that respond to faces has made the interpretation of the complete neural system for face processing challenging. The complexity of the problem is compounded by the difficulties encountered in comparing the non-human primate brain examined in single unit neurophysiology with the human brain, which is the subject of neuroimaging and neuropsychological studies.

Progress in this endeavor was made recently by Haxby and colleagues, who have integrated extensive findings from across these diverse lines of research. They proposed a *distributed neural system* for human face perception [53]. Their model emphasizes a distinction between the representation of the invariant and changeable aspects of the face, both functionally and structurally. Where function is concerned, the model posits that the invariant aspects of faces contribute to face recognition, whereas the changeable aspects of faces serve social communication functions and include eye gaze direction, facial expression, and lip movements. The proposed neural system reflects an analogous structural split. The model includes three core brain areas and four areas that extend the system to a number of related specific tasks.

For the core system, the lateral fusiform gyrus is an area activated in many neuroimaging studies of face perception. As noted previously, this region is commonly known now as the *fusiform face area* (FFA). It is the site proposed to represent information about facial identity and other categorical properties of faces. In primate studies, the homologous area is inferotemporal cortex. Though bilateral activations of this region are frequently found in human neuroimaging studies, the most consistent findings are lateralized in the right hemisphere. This is consistent with much previous neuropsychological data indicating the relative dominance of the right hemisphere in the analysis of faces.

The area Haxby et al. propose as the site of encoding for the changeable aspects of the faces is the posterior superior temporal sulcus (pSTS). Studies of single unit physiology in non-human primates and neuroimaging studies in humans indicate that this area is important for detecting gaze information, head orientation, and expression. More generally, Haxby et al. note that the perception of biological motion, including motion of the whole body, the hand, and the eyes and mouth, has been shown consistently to activate the pSTS.

The lateral inferior occipital gyri comprise the third component of the distributed neural system. Haxby et al. note that this region abuts the lateral fusiform

region ventrally and the superior temporal sulcal region dorsally, and may provide input to both of these areas. The region is proposed as a precursor area involved in the early perception of facial features, which may transit information to the lateral fusiform area and the pSTS.

In addition to the three core regions of the distributed system, four brain regions are proposed as part of an extended system for face processing. These are the intraparietal sulcus, auditory cortex, the anterior temporal area, and a set of limbic structures including the amygdala and insula. The intraparietal sulcus is involved in spatially directed attention; the auditory areas are involved in prelexical speech perception from lip movements; the anterior temporal area is involved in the retrieval of personal identity, name, and biographical information; and, the limbic structures are involved in the perception of emotion from expression. The limbic system is a part of the phylogenetically older mammalian brain that is closely tied to the survival mechanisms subserved by the emotions (e.g., fear, disgust, etc.).

Three of the four extender systems tie into specific tasks that we accomplish with different kinds of facial movement. The movements of the eyes and head direct our attention. The movements of the lips as we speak can increase the signal to noise ratio in understanding speech. Finally, the emotions are also a critical output system for facial motions. The accurate perception of emotions such as fear from the faces of others is a necessary component for adaptive social interactions.

### **Early Visual Processing and the Distributed Model**

The functional and structural deviations proposed in the distributed model between the processing of the invariant and changeable aspects of faces map easily onto what is known about the channeling of information in early vision [54]. From the retina, through the lateral geniculate nucleus of the thalamus, and up to the visual cortical regions of the occipital/temporal lobes (V1, V2, V3, V4, and V5/MT), two primary processing streams can be distinguished. These are evident both via anatomical markers such as the size and shape of the cells and the extent of their interconnections, and also via their receptive field properties.

The parvocellular stream begins with ganglion cells in the retina and is selective for the color and form, among other visual features. It projects ventrally toward the temporal cortex and ultimately toward brain regions like the fusiform area which are thought to be responsible for object and face recognition. The magnocellular stream also begins with the ganglion cells in the retina but is selective for properties of motion like speed and direction. It is also sensitive to form information but not to color. The magnocellular stream projects to parietal cortex and to regions that are close to the posterior superior temporal areas. In general, the magnocellular stream is thought to be important for processing motion information and for locating objects in the visual environment. It has also been implicated in functions that enable visual and motor coordination for action, e.g., picking up an object using vision to guide the action.

These parallel processing streams map easily onto the distributed model and are suggestive of a cortical organization that distinguishes between object properties

that may change continually and object properties that remain invariant over time. This makes for a system that can functionally divide the processing of faces into two parts: an identity analysis and an social interaction-based analysis.

### 2.3 Conclusions

Human memory for faces is characterized by robust generalization to new viewing conditions for faces that are familiar to us. Similar to many computational models, however, human abilities are far less impressive when faces are relatively new to them. Stimulus and subject factors such as the typicality of the face and the interaction between the observer’s race and the face race are strong determinants of human accuracy at the level of individual faces. These combined findings are suggestive of a system that represents faces in an image-based fashion and operates on faces in the context of a particular subject’s experience history with faces. The representation suggested is one that encodes faces relative to a global average, and which evaluates deviation from the average as an indication of the unique properties of individual faces. Although little is currently known about how facial movements affect the extraction and encoding of uniqueness information in a face, this topic is fast becoming a focus of many current studies in the literature.

The neural underpinnings of the face system are likewise complex and possibly divergent. The relatively local nature of the areas in the brain that respond to faces must be weighed against the findings that many different parts of the brain are active. The variety of tasks we perform with faces may account for the need to execute, in parallel, analyses that may be aimed at extracting qualitatively different kinds of information from faces. The tracking and interpreting of facial motions of at least three different kinds must occur while the human observer processes information about the identity and categorical status of the face. Each of these movements feeds an extended network of brain areas involved in everything from pre-lexical access of speech to lower order limbic areas that process emotion. The multifaceted nature of these areas suggests that the problem of face processing is actually comprised of many subunits that the brain may treat more or less independently.

### References

1. A.J. O’Toole, H., Abdi, K.A. Deffenbacher, D. Valentin. “Low dimensional representation of faces in high dimensions of the space.” *Journal of the Optical Society of America A*, 10, 405–410, 1993.
2. J.C. Bartlett, and J. Searcy. “Inversion and configuration of faces”. *Cognitive Psychology*, 25:281–316, 1993.
3. J. W. Tanaka, and M.J. Farah. “Parts and wholes in face recognition”. *Quarterly Journal of Psychology*, 46A:(2)225–245, 1993.
4. R.K. Yin. “Looking at upside-down faces”. *Journal of Experimental Psychology*, 81: 141–145, 69.
5. A.W. Young, D. Hellawell, and D.C. Hay. “Configurational information in face perception”. *Perception*, 16:747– 759, 1987.

6. P. Thomson. "Margaret Thatcher: A new illusion". *Perception*, 9:483–484, 1980.
7. D.M. Green, and J.A. Swets. *Signal detection theory and psychophysics*, New York, 1966
8. A.J. O'Toole, K.A. Deffenbacher, and D. Valentine. "Structural aspects of face recognition and the other-race". *Memory & Cognition*, 22: 208–224, 1994.
9. G. Rhodes. *Superportraits: Caricatures and Recognition*. Hove: The Psychology Press, 1997.
10. V. Bruce, A.W. Young. "Understanding face recognition". *British Journal of Psychology*, 77(3): 305–327, 1986.
11. G. H. Bower, M. B. Karlin. "Depth of processing pictures of faces and recognition memory." *Journal of Experimental Psychology*. 103:4, 751-757, 1974.
12. A.M. Burton, V. Bruce, N. Dench. "What's the difference between men and women: Evidence from facial measurement". *Perception*, 22:(2)153–176, 1993.
13. H. Abdi, D. Valentine, B. Edelman, and A.J. O'Toole. "More about the difference between men and women: evidence from linear neural networks and the principal-component approach". *Perception*, 24:539–562, 1995.
14. A.J. O'Toole, T. Vetter, N.F. Troje, and H.H. Buelthoff. "Sex classification is better with three-dimensional head structure than with image intensity information". *Perception*, 26:75–84, 1997.
15. H. A. Wild, S.E. Barrett, M.J Spence, A.J. O'Toole, Y. Cheng, and J. Brooke. "Recognition and categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues". *Journal of Experimental Child Psychology*, 77: 269–291, 2000.
16. D.M. Burt, and D.I. Perrett. "Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information". In *Proceedings of the Royal Society London B*, 259:137–143, 1995.
17. A.J. O'Toole, T. Vetter, H. Volz, and E.M. Salter. "Three-dimensional caricatures of human heads: Distinctiveness and the perception of facial age". *Perception*, 26: 719–732, 1997.
18. P.J. Ekman, and W.V. Friesen. *The facial action coding system: A technique for the measurement of facial movement*. San Francisco: Consulting Psychology Press, 1978.
19. A.J. O'Toole, D. Roark, H. Abdi. "Recognition of moving faces: A Psychological and neural perspective". *Trends in Cognitive Sciences*, 6:261–266, 2002.
20. L. Light, F. Kayra-Stuart, and S. Hollander. "Recognition memory for typical and unusual faces". *Journal of Experimental Psychology: Human Learning and Memory*, 5:212–228, 1979.
21. T. Valentine. "A unified account of the effects of distinctiveness, inversion, and race in face recognition". *Quarterly Journal of Experimental Psychology*, 43A:161–204, 1991.
22. V. Blanz, T. Vetter. "A morphable model for the synthesis of 3D faces". In *SIG-GRAPH'99 Proceedings*, ACM: Computer Society Press, 187–194, 1999.
23. V. Blanz, A.J. O'Toole, T. Vetter, H.A. Wild "On the other side of the mean: The perception of dissimilarity in human faces". *Perception*, 29, 885–891, 2000.
24. D. Leopold, A.J. O'Toole, T. Vetter, V. Blanz, "Prototype-referenced shape encoding revealed by high-level aftereffects". *Nature Neuroscience*, 4, 89-94, 2001.
25. I. Craw, P. Cameron. "Parameterizing images for recognition and reconstruction". In *p. mowforth (ed.) Proceedings of the British Machine Vision Conference London: Springer-Verlag*, 1991.
26. A. Lanitis, C.J. Taylor, T.F. Cootes. "Automatic interpretation and coding of face imaging using flexible models". *IEEE Transactions Pat. Anal. Mach. Intell.*, 19: 743, 1997.

27. M. Turk, and A. Pentland. “Eigenfaces for recognition”. *Journal of Cognitive Neuroscience*, (3): 71–86, 1991.
28. R.S. Malpass, and J. Kravitz. “Recognition for faces of own and other race faces”. *Journal of Personality and Social Psychology*, 13:330–334, 1969.
29. D.S. Lindsay, P.C. Jack, and M.A. Christian. “Other-race face perception”. *Journal of Applied Psychology*, 76:587–589, 1991.
30. D.S. Furl, P. J. Phillips, A.J. O’Toole. “Face recognition algorithms as models of the other-race effect”. *Cognitive Science*, 96:1–19, 2002.
31. M. Riesenhuber, T. Poggio. “Models of object recognition”. *Nature Neuroscience Supplement*, 3:1199–1204, 2000.
32. I. Biederman, P. Gerhardstein. “Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance”. *J. Exp. Psychol.: Hum. Percept. Perform.*, 19:1162–1183, 1993.
33. D. Marr, *Vision*, San Francisco : Freeman, 1982.
34. T. Poggio, S. Edelman. “A network that learns to recognize 3D objects”. *Nature*, 343 :263–266, 1991.
35. S. Ullman, *High-level vision*, Cambridge, MA: MIT Press, 1996.
36. A.M. Burton, V. Bruce, P.J.B. Hancock. “From pixels to people: a model of familiar face recognition”. *Cognitive Science*, 23:1–31, 1999.
37. P.J.B. Hancock, V. Bruce, A.M. Burton “Recognition of unfamiliar faces”. *Trends in Cognitive Sciences*, 4:(9)263–266, 1991.
38. P.J.B. Hancock, V. Bruce, A.M. Burton “Recognition of unfamiliar faces”. *Trends in Cognitive Sciences*, 4:(9)263–266, 1991.
39. Y. Moses, S. Edelman, S. Ullman “Generalization to novel images in upright and inverted faces”. *Perception*, 25:(4)443–461, 1996.
40. N.F. Troje, H.H. Bülthoff. “Face recognition under varying pose: The role of texture and shape”. *Vision Research*, 36:1761–1771, 1996.
41. S. Edelman, H.H. Bülthoff. “Orientation dependence in the recognition of familiar and novel views of three-dimensional objects.”. *Vision Research*, 32(12):2385–2400, 1992.
42. W. L. Braje, D.J. Kersten, M.J. Tarr, N.F. Troje “Illumination effects in face recognition.”. *Psychobiology*, 26:371–380, 1999.
43. H. Hill, V. Bruce. “Effects of lighting on the perception of facial surface”. *J. Exp. Psychol. : Hum. Percept. Perform.*, 4:(9)263–266, 1991.
44. A.M. Burton, S. Wilson, M. Cowan, V. Bruce. “Face recognition in poor-quality video,” *Psychological Science*, 10:243–248, 1999.
45. J. Kurucz, J. Feldmar. “Prosopo-affective agnosia as a symptom of cerebral organic brain disease,” *Journal of the American Geriatrics Society*, 27:91–95, 1979.
46. J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Shouten, J.L. Pietrini. “Distributed and overlapping representations of faces and objects in ventral temporal cortex”. *Science*, 293:2425–2430, 2001.
47. M. Spiridon, N. Kanwisher. “How distributed is visual category information in human occipito-temporal cortex? An fMRI study”. *Neuron*, 35:1157, 1991.
48. K. Tanaka. “Neuronal mechanisms of object recognition”. *Science*, 262:685–688, 1991.
49. N.K. Logothetis, J. Pauls, H.H. Bülthoff, T. Poggio. “Shape representation in the inferior temporal cortex of monkeys”. *Current Biology*, 5:552–563, 1991.
50. D. Perrett, J. Hietanen, M. Oram, P. Benson. “Organization and function of cells responsive to faces in temporal cortex”. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci*, 335: 23–30, 1992.

51. I. Gauthier, M.J. Tarr, A.W. Anderson, P. Skudlarski, J.C. Gore. "Activation of the middle fusiform face area increases with expertise recognizing novel objects". *Nature Neuroscience*, 2:568–, 1999.
52. N. Kanwisher, J. McDermott, M. Chun. "The fusiform face area: a module in human extrastriate cortex specialized for face perception" *J. Neurosci.*, 17 :4302-4311, 1997.
53. J.V. Haxby, E.A. Hoffman, and M.I. Gobbini. "The distributed human neural system for face perception". *Trends in Cognitive Sciences*, 20(6):223–233, 2000.
54. W.H. Merigan. "P and M pathway specialization in the macaque" In A. Valberg and B.B. Lee (Eds.). *From Pigments to Perception*. Plenum Press, 117–125, 1991.