CHAPTER 12
RECOGNISING FACES

CHAPTER SUMMARY

Introduction 314
Structural descriptions 314
Some theories of visual object recognition 315
The nature of face processing 317
Expertise and face recognition 318
  Face recognition: configural processing 320
  Representation of faces in the brain: norm-based coding 321
Average faces, caricatures and anti-faces 323
Face aftereffects and face-specific mechanisms 327
Neural basis of face identification 329
Controversies in studies of face recognition 335
Overview of studies of face recognition 336
Summary 336
Test yourself questions 337
Suggestions for further reading 337
PERCEIVING THE WORLD AROUND US

INTRODUCTION

Humans are social animals, whose way of living requires us to compete and cooperate with other members of our species. To do that, we need to identify other individuals quickly and accurately. Although, as we shall see, some individuals have great difficulty in recognising people from their faces, for almost all of us face recognition is usually effortless and reliable. Life would be very different if we usually misidentified spouses and lovers, friends and enemies. In this chapter, the following questions will be addressed:

- Why are faces easier to recognise than other objects only when presented the right way up?
- How are faces represented in the brain?
- Why do some neurological patients need to see more than their spouse’s faces before they can recognise them?
- Do we recognise faces by comparing them with some form of stored average face?
- What is ‘face space’?
- Why are caricature faces so easy to recognise?

STRUCTURAL DESCRIPTIONS

In an influential review, Bruce and Young (1986) identified a number of processes which occur in the visual system when we recognise a face. An important early stage is the generation of a structural description or code. This description has to be immune to changes in the appearance of a face produced by changes in lighting, make-up, age, expression, and hairstyle since we can recognise individuals despite such changes. The exact nature of this description is not clear, but its immunity to change may be achieved partly by the weighting given to particular regions of the face, such as the eyes, mouth, and nose (internal features). It turns out that, in studies of familiar faces, these features when presented alone lead to better recognition than do external features (ears, hairline, hairstyle) when presented alone (Ellis et al., 1979). For unfamiliar faces, however, both types of feature are equally important in recognition. External features are more likely to change with fashion or age than internal features. Hair may fall out, and/or be worn longer or shorter, and ears grow larger with age, but eyes and noses, and their relative positions, stay much the same, and so a description based on them is likely to be more reliable.

Some attempts to model face recognition have not paid much attention to the details of the structural description of faces, but rather have emphasised the role of other information in identifying a person (e.g. Burton et al., 1990). This includes semantic information (e.g.
RECOGNISING FACES

‘Queen Elizabeth is married to a man named Philip’), and idiosyncrasies of speech. Though clearly important for understanding human cognition, such work is beyond the scope of this book. Instead, this chapter covers some issues which surround the notion of a structural description of a face, namely how face recognition is related to the recognition of other objects, how, if at all, face recognition is special, and whether there is a brain region specialised for face recognition.

SOME THEORIES OF VISUAL OBJECT RECOGNITION

It is clear that, when we recognise an object, some representation based on incoming sensory information is matched to some stored representation of the object. In principle, these representations could be somewhat abstract, rather as a written list of an object’s properties is a non-visual representation of the object. However, in practice, theories of recognition by humans have described the representation in terms of the forms of (parts of) objects and how they are put together.

The image on the retina is represented in various ways in the nervous system. David Marr (1982) distinguished three such representations. In the ‘primal sketch’, areas of similar intensity in the retinal image are grouped together. This might be done in Area V4 — see Chapter 3. The ‘viewer-centred’ representation describes the layout of surfaces relative to the position of the viewer. At least some of this may occur in V4 also. The ‘object-centred’ representation (which is particularly relevant in this context) is a viewpoint-independent representation of the three-dimensional scene. Marr described objects in the object-centred representation in terms of ‘generalised cylinders’. Thus a person could be thought of as composed of a number of cylinders (calves, thighs, forearms, trunk, etc.), in a particular arrangement. An ape would be composed of slightly different cylinders (longer arms, etc.) in a similar arrangement, whereas, in the case of an elephant, the cylinders would be different, as would their arrangement. One important aspect of this scheme is that it incorporates models at different spatial scales. Thus in the model representing a human, the forearm and hand could be represented by a single cylinder, whereas a model representing individual fingers or parts of fingers could be selected for the more detailed recognition of a hand (see Figure 12.1).

Although Marr is justly famous for his pioneering work in the computational modelling of vision, it is only fair to point out that these are ideas rather than working computer programs which can process natural scenes, and other additional processes which have not yet been made explicit may be needed to make them work.

A development based on Marr’s idea is that of Recognition by Components (Biederman, 1987). The components in this theory are three-dimensional shapes (such as spheres and pyramids, as well
PERCEIVING THE WORLD AROUND US

Figure 12.1 Representation at different spatial scales in visual object recognition. Recognising from a distance that something is a human being requires only a rather coarse representation (as in the left-hand box), whereas recognising that something is a human hand in a particular orientation requires a much more detailed representation (right-hand box). Each representation is made up of ‘generalised cylinders’.

Source: Marr D foreword by Shimon Ullman, afterward by Tomaso Poggio (2010) Vision: a computational investigation into the human representation and process of visual information. Figure 5.3 © Massachusetts Institute of Technology, by permission of The MIT Press.

1. Cross-section: straight vs. curved

2. Axis: straight vs. curved

3. Size of cross-section: Constant (parallel sides) vs. expand vs. expand & contract vs. contract & expand

4. Termination of geon when non-parallel: truncated vs. pointed vs. rounded

Figure 12.2 Examples of geons produced by sweeping circular and square cross-sections through various trajectories. Example 1 shows the results of sweeping a square and a curved cross-section through a straight axis, giving a brick and a cylinder. Example 2 shows the results of varying the axis from straight to curved. The other examples show the effects of more complex manipulations.
RECOGNISING FACES

as cylinders) known as ‘geons’ (geometrical icons). Some examples, and how to generate them, are shown in Figure 12.2. Again, objects are represented by a collection of geons, together with a description of their orientations and spatial relationships. Although this has been successfully implemented as a neural network (Hummel and Biederman, 1992), the inputs to the network were line drawings. Biederman has suggested that the human visual system converts its input to a line-drawing-like representation at an early stage of processing, but several lines of evidence are not consistent with this. For example, Attwood et al. (2001) showed that visual search for line-drawn targets was slower than that for targets defined by shading, even after extended practice. This would be unexpected according to the theory, since the line drawings would require no conversion.

A version of the suggestions of both Marr and of Biederman may account for face recognition, in the sense of recognising that something is a face rather than a dog or a car. Thus the central area of a face might be thought of as two partially occluded spheres (eyes) above a pyramid (nose) which is in turn above a horizontal cylinder (mouth). This idea becomes somewhat strained, however, when we consider facial identification, the further stage of recognising that the face is that of a particular person. This individual, and most humans, have two eyes, one nose, and one mouth in a similar configuration. It is small differences in these features (‘prominent blue eyes’, ‘a snub nose’, and ‘thin lips’) and in the spatial relationships between them (‘close-set eyes’) that identify a particular individual. To recognise that something is a face seems to require that the geons be rather vaguely defined (so that, say, a pyramid represents snub, roman and pointed noses) whereas identifying the face requires that geons be very specific. Similarly, specifying the spatial relationships between components should be vague at the generic level, but precise at the individual level. It is not clear how these two conflicting requirements could be accommodated by the present version of the Recognition by Components theory. Indeed, Biederman and Kalocsai (1997) have argued that the ‘individuation’ (or identification) of faces and the classification of objects are very different processes, with the geon-based model being applicable only to the latter. Rather, they suggest, facial identification requires information about sizes of and distances between features in a holistic representation of the facial surface, which seems like a fuller description of Bruce and Young’s structural code.

THE NATURE OF FACE PROCESSING

The upsurge in interest in face recognition was sparked more than 40 years ago by a demonstration of what is now known as the inversion effect (Yin, 1969). If people study a series of pictures of faces, and of other objects such as houses or aeroplanes, they tend to perform better at recognising the faces than they do the other objects. But if the pictures are inverted, performance on the faces is worse than on the other objects (see Figure 12.3). Yin concluded that face recognition is special since the process is worse affected by inversion of the stimuli than is the recognition of other objects.
EXPERTISE AND FACE RECOGNITION

Others have questioned Yin’s conclusion. One possibility is that face recognition is not necessarily different in nature from other forms of recognition, but that we are experts at it. Because the process is ‘over-learned’, we are better at recognising upright faces than upright houses, but are more vulnerable when faces are presented in an unusual orientation. This idea can be tested by studying individuals who are expert at recognising other types of object (see Key study), either naturally occurring, or specially designed for experimental purposes (e.g. ‘Greebles’).

KEY STUDY

EXPERTISE IN OBJECT RECOGNITION

Some occupations (e.g. chicken sexing) require considerable expertise in recognition. In one study, dog breeders, experts at judging the finer points of canine beauty, studied photographs of examples of the breed in which they specialised. In a later recognition test, they were better at recognising those individuals than were non-experts, but showed a larger inversion effect when the stimuli were presented upside down (Diamond and Carey, 1986). This result certainly supports the expertise hypothesis, but it could be argued that results reflected some innate ability, not acquired expertise. What we really need to know is how the experts would have performed before they developed expertise, as well as how they did afterwards. Gauthier and colleagues (1998) devised a set of android-like creatures, which they called ‘Greebles’. An example is shown in Figure 12.4. A Greeble could belong to one of several families, and within a family each Greeble was identified by particular attributes. When participants...
RECOGNISING FACES

initially took part in recognition tests of Greebles, they performed moderately, as they might on houses, and the inversion effect was small. However, with repeated training, recognition performance with upright Greebles improved, and the inversion effect with upside-down Greebles increased. This is a clear demonstration that expertise is involved in the inversion effect, because we know about performance before training and how it changed as expertise developed.

Although Greebles are not human faces, they might engage to some extent the mechanisms involved in recognising persons. They appear to have a kind of head and body, together with various protruberances, just as humans do. Although face recognition is perhaps the major route to person identification, there is evidence that other aspects of the body are involved. It turns out that there is an inversion effect for bodies as well as faces (Reed et al., 2003), suggesting that we have developed expertise in body as well as in face recognition.

Other evidence to support the role of the body in person identification comes from studies of biological motion (see Chapter 5), which have shown that the walker’s gender can be perceived in such displays (Mather and Murdoch, 1994), and even the identity of familiar individuals (Stevenage et al., 1999). These effects are profoundly disrupted if the movie is shown upside down (Pavlova and Sokolov, 2000).

Perhaps, then, the expertise developed by training on Greebles is mediated by processes underlying normal person identification. Can a similar effect of expertise be developed for recognition of stimuli which are nothing to do with faces? Husk et al. (2007) generated a series of pictures of the fronts of houses (an example is provided in Figure 12.5). Initially, their observers performed moderately on tests of recognition and showed a small inversion effect. However, they improved with training to levels associated with typical performance in face recognition tasks, and then showed a large inversion effect. Again, though, one could argue that the houses appear to have face-like qualities, with the upper windows representing eyes, and the door the nose. In a second experiment,
Hussain et al. (2009) ruled out this idea. They produced a series of stimuli composed of different types of texture (see Figure 12.6 for an example). Initially, recognition was poor, and inverting the images had little additional effect. However, after training, performance improved substantially, and a large inversion effect was found. It is hard to argue that recognition of random textures is closely related to the processes of face recognition, so this experiment provides good evidence that recognition expertise can be developed in processes other than those of the recognition of faces or other familiar objects.

**KEY POINTS**

- Faces are recognised better than other objects, and their recognition is impaired more by turning them upside down (the inversion effect). The large inversion effect for faces has been taken as evidence that face recognition is special
- Large inversion effects have also been found for recognition of dogs in dog breeders, suggesting that they reflect an effect of expertise not specific to faces
- Large inversion effects were found after training in recognition of Greebles (android-like objects), houses, and textures. The latter effect especially suggests that the effects of expertise are not unique to faces

**Face recognition: configural processing**

Exactly why should inversion affect the recognition of faces (and other over-learned stimuli) so severely? The general view is that faces are not processed as a collection of individual features but as a configuration, and it is this configural processing which is particularly susceptible to inversion. What is needed to support this idea is evidence for configural processing which is independent of the effects of inversion. An example of this is the ‘composite face effect’ of Young and colleagues (1987), who studied the ease with which individual features of faces could be recognised. In some of their conditions, faces were separated into top and bottom halves and, in others, the bottom half of one face was combined with the top half of another face, to form a composite face (see Figure 12.7 for an example). The authors found that it was much easier to identify a feature in the original face, or in half of the original face, than it was in a composite face. They suggested that configural processing acted to combine the upper and lower halves of different faces into a new identity, which put individual features into a novel context and so hindered their identification. However, they also showed that inverting the composite faces or misaligning their top and bottom halves largely removed this disadvantage in identifying individual features. These experiments show that normal face processing is configural, cannot be prevented even when the task would benefit from feature by feature processing, and does not operate when faces are inverted.
Can we say more about the details of configural processing? Maurer et al. (2002) distinguished three separable aspects of configural processing of faces:

1. Perception of first-order relationships. Seeing that two features, which could be eyes, are located above what could be a nose, which in turn lies above something that could be a mouth. This aspect of perception could be carried out by, for example, the Recognition by Components model outlined above.

2. Holistic processing. What Maurer et al. describe as the ‘gluing together’ of eyes, nose and mouth into a ‘gestalt’ or whole (a process which seems to underlie the Composite Face Effect).

3. Perceiving second-order relationships. Seeing the distances between different features (‘close-set eyes’, etc.), a process which is likely to be important in identifying individual faces, and similar to that suggested by Biederman and Kalocsai (1997), mentioned earlier.

It turns out that these three types of processing are affected differently by inversion. Perception of first-order relationships (seeing that an object is a face) seems to be affected only to a small extent, if at all, whereas the perception of second-order relationships (identifying faces which are distinguished by the separation between their features) is markedly impaired by inversion (Freire et al., 2000). As we have seen, the Composite Face Effect (a hallmark of configural processing) does not occur for inverted faces.

**Representation of faces in the brain: norm-based coding**

How are faces represented in the brain? One possibility is that we have a face ‘prototype’ or ‘norm’ stored in our brains. Such a prototype is of a typical face, built up as some kind
of average of all the faces which we have seen, biased towards the characteristics of more recently seen faces. Exactly how the features in this prototype are coded is uncertain, but we are thought to identify individual faces by deviations from it (Leopold et al., 2001, 2005). Notice that this idea fits well with the way we describe individual faces. Terms such as ‘snub nose’ or ‘bushy eyebrows’ only have meaning if the hearer has an idea of the length of the average nose or the amount of hair in the average eyebrow. It also fits well with the idea that individual faces are identified, at least in part, by their second-order relationships, with the addition that distances between features are coded in terms of their deviations from the dimensions of the average face.

Many journal articles on face perception refer to ‘face space’, a graph on which any point uniquely identifies a particular face. An example, which can be drawn on paper, is shown in Figure 12.8. The graph has three axes, or dimensions. The x-axis specifies length of nose, from ultra-snub to pinnchioesque, the y-axis colour of skin, from jet black to snow white, and the z-axis separation of the eyes, from almost touching to far apart. The axes are at right angles to each other, or orthogonal, and so the dimensions are independent of each other. This means that points representing two individuals with identical snubbish noses and closeish eye separation would occupy the same positions on the x–z plane, but the difference in their skin colour (one white – brown point, and one bronzed – blue point) would be

Figure 12.8  3D graph illustrating Face Space. The three axes represent length of nose, colour of skin, and separation of eyes. These three dimensions are independent. The graph shows the points in Face Space occupied by three different faces, and by the average face. See text for more detail.
RECOGNISING FACES

specified by a difference of their points on the y-axis, which can be adjusted without interfering with the position of those points on the x and z dimensions. In contrast, a black individual, with a long nose and eyes far apart – black point – would occupy very different positions on the x and z dimensions, as well as on the y dimension. Of course, individual faces differ in more than three ways, but it is not possible to draw graphs in our 3D world with more than three dimensions.

Conveniently, however, mathematicians can conceive of graphs (which cannot be drawn on a flat surface) with many more dimensions, all orthogonal, and so all independent. This collection of dimensions defines a space in which any face can be represented by a single point, corresponding to its value on all the relevant dimensions. It is uncertain at present exactly how many dimensions are ‘relevant’ to encoding a face, or their relative importance, although, for example, length of nose seems likely to be more important than length of eye-lashes, in general. Note that somewhere in the middle of face space lies a point which defines the average face. In the 3D graph above, it is the green point, defining a face with a nose of medium length, bronzed skin and eyes a medium distance apart. Such a face has been called the norm, or prototype face.

AVERAGE FACES, CARICATURES AND ANTI-FACES

Although the concept of the average face has only recently become important in face research, it goes back more than a century. Francis Galton (1879) used photographic techniques to merge images of several component faces to produce a composite face, of which he remarks: ‘All composites are better looking than their components, because the averaged portrait of many persons is free from the irregularities that variously blemish the looks of each of them’ (1879: 135). This conclusion has been rediscovered more recently (see Langlois and Roggman, 1990). Modern researchers merge images using computers, but the underlying notion is the same.

Figure 12.9 shows an example of such a face in which the sizes and positions of the various features are the means of the sizes and positions of those features in a large number of individuals. In principle, such a face could be representative of the entire human race, if the average were taken over all races, ages and genders. The face in Figure 12.9 is that of the average of a sub-set of the human race, namely young Caucasian males. Cartoonists caricature well-known individuals by exaggerating certain features, such as the size and shape of their noses, or the extent to which their ears stick out. Again, this technique has a long history, and is often used unkindly to portray individuals such as politicians. Indeed, it is a mark of fame (of a kind) that a caricature is enough to identify someone in a newspaper cartoon, whereas a less well-known person might have
a label somewhere on their caricature. Face researchers produce caricatures by measuring (with suitable software) the differences in size and position of features from those in the average face, and then increasing them (or some of them). Because average faces and caricatures have been around for at least a century, face researchers might be thought of as having used high technology simply to re-invent the wheel. However, one novel development, facilitated by computer technology, is the ‘anti-face’, which is the reverse of a caricature. Anti-faces are produced by first measuring the differences between the features of an individual face and those of the average face (as in the first stage of caricaturing), but then subtracting some fraction of the values from those of the average face. So the anti-face of Mick Jagger would have very thin lips, and that of Prince Charles would have ears which were closely moulded to the side of the head.

One line of evidence which supports the face prototype idea comes from studies of face aftereffects. In Chapter 1, we saw that staring for a while at simple stimuli, such as lines of a particular orientation, could distort the subsequent perception of similar but not identical lines. So after one has stared at lines tilted anticlockwise from vertical, vertical lines appear to be tilted clockwise from vertical.

Webster and Maclin (1999) showed that similar aftereffects could arise in the visual perception of faces. They systematically distorted the internal features of faces with neutral expressions, by expanding or contracting them, vertically and/or horizontally, relative to the midpoint of the nose. Points close to the midpoint were moved more than those further away (to be more exact, the shifts were weighted by a Gaussian window). This produced a set of images whose distortion could be quantified. The authors found that, after adapting to a distorted face, the normal face needed to be distorted in the same direction as during adaptation to appear normal. In other words, the effect of adaptation was to make the normal face appear distorted in the opposite direction to that during adaptation. You can try this for yourself, using the faces in Figure 12.10. First, inspect the original face (in the centre of the third row), noting its general characteristics. Now, stare for a minute or so at the face at the bottom left of the array. Then look back at the central face — it should now appear distorted, and look more like the face at the top right. If you stare at the top right face for a while, the normal face will appear more like the bottom left face. So, after adapting to a face in which some features are expanded, those features in a normal face appear contracted, and vice versa. Webster and Maclin noted that the aftereffects were asymmetrical, so that adapting to a normal face had little effect on the perception of distorted faces. The authors suggested that the representation of the normal face had been built up by the frequent viewing of faces in everyday life, and so would be little affected by viewing other normal faces. However, it would be affected by viewing faces which differed markedly from normal.

In subsequent work, Webster and his colleagues (2004) showed that several judgements often made about faces (gender, race, and the emotions which they express) could be affected by adaptation. To examine the perception of gender, they selected two faces, one always judged to be male, the other always judged to be female. With morphing software, they produced a series of images in which the characteristics of the male face were progressively changed to make it more like the female face. The face in the middle of the series was androgynous: it was hard to judge whether it was male or female. However, after adaptation to the male face, the androgynous face was judged to be female, and after adaptation to the female face, it was judged to be male. Similar
Figure 12.10 Facial images in a study by Webster and Maclin (1999). The original face is in the centre of the third row. Faces to the right of the original have their features expanded horizontally, so that, for example, the nose becomes progressively wider. To the left of centre, internal features are progressively more compressed horizontally, so that the nose becomes narrower. Faces above the third row have their internal features progressively expanded vertically, so that, for example, the nose becomes longer, and internal features in faces below the third row are contracted vertically, so that the nose becomes shorter.


effects were found for race, using a series of morphs between Caucasian and Japanese, and for emotion, for example, with morphs between happy and angry. Thus, the stored face prototype may be neutral—in expression, race and emotional content.

The prototype is also likely to be of an upright face with normal contrast polarity, so that the relative lightnesses of areas which are normally lighter (the cheeks, say) and those which are normally darker (the pupils of the eyes, say) are preserved. We have already seen that inverting faces during recognition tests markedly reduces accuracy. Similarly, reversing the contrast of faces (as
in photographic negatives) impairs recognition of faces more than of other objects (Kemp et al., 1996; Nederhouser et al., 2007), as suggested by Figure 12.11.

Adaptation to upright and inverted faces, and to faces with normal and reversed contrast, can produce opposite aftereffects. Thus, after adaptation to, say, upright expanded faces, interspersed with adaptation of inverted contracted faces, upright faces appear contracted and inverted faces appear expanded (Rhodes et al., 2004). This finding suggests that upright and inverted faces engage different neural processes in the brain (as indeed we might suppose from the large differences in accuracy of recognition of upright and inverted faces). It is also an example of contingent adaptation, in which the direction of the aftereffect depends on another stimulus property which is consistently present during adaptation (upright faces with expansion, and inverted with contraction). Such effects are reminiscent of contingent aftereffects, such as the McCollough Effect (McCollough, 1965). To obtain the McCollough Effect participants viewed for 2–4 minutes vertical orange and black stripes, alternating say every 15 seconds with identical but horizontal black and blue stripes. One half of the test display consisted of vertical black and white stripes and the other half of horizontal black and white stripes. The vertical stripes appear tinged with blue/green, and the horizontal stripes tinged with orange. Thus the stripes appeared to be of the complementary colour to the physical colour paired with that orientation during adaptation.

Face aftereffects are certainly robust phenomena, showing clearly that adapting to a distorted face can influence the perception of subsequently viewed faces, but do they really support the notion of norm-based coding? Participants of Leopold et al. (2001) were trained to identify target faces for which anti-faces had been constructed. They were then asked to identify each target face, as before, but after adaptation to an anti-face, which could be either that of the target or that of another face in the target set. Adaptation to the target anti-face improved identification, whereas adaptation to another anti-face did not. The authors concluded that this supported norm-based coding because adaptation to the target anti-face would shift the target further from the norm than would adaptation to another anti-face, and so make its identification easier.

Rhodes and Jeffrey (2006) questioned how well this result supported the conclusion. Rather than being selective for the opposite facial identity, and so supporting norm-based coding, they suggested that the effects of adaptation might be very general and so lead to the distortion of any test face. They pointed out that the other anti-faces in the set from which the adapting face on any trial was drawn were likely to be more similar to the target face, simply because the anti-face, by
definition, had been constructed to be very different from the target face. Because aftereffects can be reduced if the differences between the adapting and test stimuli (‘perceptual contrast’) are too small, then any generalised effect could have the greatest effects on the target face, and so lead to enhanced identification. What is needed to test this idea are adapting faces whose differences from the norm are of the same size as those of the anti-face, but which are of different identities. If perceptual contrast determines the size of the aftereffects, then all such faces should yield face aftereffects of a similar size. If aftereffect size is determined by the change in the prototype, then the target matched to the adapting anti-face should give the largest aftereffect. Rhodes and Jeffrey obtained difference ratings for a range of faces, and so were able to construct a range of anti-faces whose differences from the targets were of similar magnitude, but in which only one anti-face matched each target face. They found that the aftereffects of adapting to these anti-faces were largest when the target matched the anti-face, thus suggesting that differences in perceptual contrast cannot explain the findings of Leopold et al. (2001), and supporting the idea of norm-based coding of faces.

**THINKING ABOUT RESEARCH**

**CAR AFTEREFFECTS?**

Face aftereffects suggest that faces may be encoded as deviations from a norm or prototype. Might other familiar objects be encoded in a similar way? With a very small number of exceptions, a car is essentially a rectangular box with a wheel at each of its lower corners. However, the ratio between length and height can vary significantly between different models. For example, a Jaguar or Aston Martin sports car is many times longer than it is high, whereas the length of a Range Rover or similar 4 × 4 may be much less than twice its height. The length:height ratio of a family saloon lies somewhere between the two (and might approximate a prototype, if such a thing exists for cars). Design a study to demonstrate car aftereffects. How could you measure any such effects? With reference to the literature on face recognition, include conditions which might rule out adaptation to local features as the basis for any aftereffects.

**FACE AFTEREFFECTS AND FACE-SPECIFIC MECHANISMS**

So far we have accepted that face aftereffects are just that – whatever the neural processes which underlie face perception, it is these that are affected by adaptation to distorted faces. However, other aftereffects, such as those of tilt or motion, are thought to reflect adaptation in
PERCEIVING THE WORLD AROUND US

mechanisms which process simple physical characteristics of the retinal image. Perhaps these low-level mechanisms adapt, say, to the curve of an eyebrow or the width of a nose in a distorted face. When a neutral test face with less curved eyebrows and narrower nose is subsequently viewed, these lower level mechanisms would signal even straighter eyebrows and an even narrower nose. Even if face-sensitive mechanisms themselves were completely unaffected by adaptation, the change in these inputs could cause a shift in the perception of the test face. On this view, facial images are simply rather complex devices for demonstrating examples of low-level visual coding.

How could one decide between low- and high-level coding as explanations for face aftereffects? There are similarities between the temporal characteristics of low-level and face aftereffects. Leopold et al. (2005) adapted their participants to an anti-face, and then had them identify the person depicted in the average face. The probability of reporting the target face from which the anti-face had been constructed rose with adapting time, and fell with duration of the test presentation, as is found for other aftereffects, such as the tilt aftereffect (Magnussen and Johnsen, 1986; Harris and Calvert, 1989). Had the build-up and decay of face aftereffects been very different from that of the tilt aftereffect, one could argue that the underlying processes must be very different, but this is not what was found.

Attempts to dissociate high- and low-level aftereffects have concentrated on transfer – how different from the adapting face the test face can be, and yet support an aftereffect. Leopold et al. (2001) changed the size of their face stimuli between the adapting and test conditions, thus changing their low-level characteristics. Nevertheless, they found large aftereffects. Rhodes et al. (2004) repeated the experiment in which they adapted upright and inverted faces to opposite distortions, but in this study the sizes of the adapt and test faces differed by a factor of more than two. Because individual features would be larger and more widely spaced in the larger images, they would be likely to show reduced transfer of the aftereffects (the authors argued), if these were produced by low-level mechanisms. Despite this, they found aftereffects for upright faces which were similar in magnitude to those found when the adapt and test faces were the same size. They also found aftereffects for inverted faces, but these were smaller than when the face size was not varied. They concluded that low-level mechanisms were not contributing much, if at all, to the aftereffects with upright faces, but were contributing to those with inverted faces.

**KEY POINTS**

- Facial expressions are processed as a configuration, as shown by the Composite Face Effect for expressions.
- Individual faces may be represented in the brain as deviations from a prototype, or norm, built up as an average of previously seen faces, with a bias towards more recently seen faces.
RECOGNISING FACES

- A face can be represented by its position in face space, a set of orthogonal dimensions encoding features such as length of nose, skin tone, etc.
- Evidence for norm-based coding comes from face aftereffects, the distortions of a normal face which follow adaptation to a face distorted in the opposite direction
- Face aftereffects for upright faces reflect adaptation in face-specific processes, not changes in the inputs to those processes

THE NEURAL BASIS OF FACE IDENTIFICATION

It has been suggested for more than 60 years that there is a centre in the brain specialised for identification of familiar faces. The initial basis for this suggestion was the neurological symptom of prosopagnosia, from the Greek for ‘not knowing faces’ (Bodamer, 1947). It was first noticed in individuals who had suffered brain damage to regions of the temporal lobe after an accident or stroke, who then could not recognise their spouse or other members of their family from their faces alone. These patients could, however, recognise their relatives from their voices or usual clothing.

Prosopagnosia certainly suggests that face recognition can be disrupted by brain damage to parts of the temporal lobe. However, some researchers have gone beyond this to claim that there is a region in the human brain specialised for face recognition alone. One issue in evaluating this is the selectivity of the impairments exhibited by particular individuals. According to De Renzi (1986), some patients do show an impairment purely for faces, and one patient performed better on recognising inverted than upright faces (Farah et al., 1995), which is the opposite of normal performance. But such claims raise the question of how rigorously the patients were tested for other impairments. For example, de Gelder et al. (1998) found a similar reversed inversion effect in another prosopagnosic patient, but for shoes. Such results have been thought to cast doubt on the idea of a specialised neural site for face recognition, but by themselves they cannot conclusively settle the question. Strokes do not respect functional boundaries in the brain. If there is a genuine face area, there must be areas which mediate the recognition of other objects, perhaps in neighbouring regions of the brain. If a stroke damaged this wider region, then recognition of many classes of objects, including faces, would be impaired, but this would not rule out the existence of a specialised face area. To do this in prosopagnosia would require the intensive study of different types of recognition in a group of patients, and include detailed measurements and localisation of their brain lesions. Only then could one hope to identify a particular area which was lesioned in patients with a range of impairments, as well as prosopagnosia, and associated with impairments in face recognition.
One difficulty in carrying out such studies is that of recruiting a sufficiently large number of suitable participants.

Historically, the second type of study to throw light on the neural basis of face processing came not from neuropsychological studies of humans, but from neurophysiological recordings from single cells in the monkey brain (Gross et al., 1969). Cells which respond to faces have been found in the amygdala (Rolls, 1992), a region known to be involved in processing emotional expressions, the ventral striatum (Williams et al., 1993), which receives inputs from the amygdala, and the frontal lobe (O Scalaidhe et al., 1997). However, the face cells which have been most extensively studied are found in and around the superior temporal sulcus (STS), an area in the temporal lobe (see Figure 12.12). One problem with interpreting neurophysiological recordings from such cells is the limitation on the set of stimuli which is shown to the animal before the cell or the experimenter stops working. Clearly, showing that a cell responds to faces does not tell us much if the cell responds well to many other objects also. Perrett and colleagues (1994) reviewed the properties of ‘face’ cells in the STS. The responses of most to faces were at least twice (and in some cases 10 times) as great as those to other three-dimensional objects, suggesting that they are specifically tuned to faces. Not all face cells in the STS are sensitive to different facial identities. Those that are sensitive are found largely in an area including the lower lip of STS, in areas TEx and TEM (see Figure 12.12). Hasselmo et al. (1989) presented their monkeys with nine pictures of monkey faces while recording from face cells. The pictures were of three different monkeys, each exhibiting three different expressions (calm, mouth slightly open, indicating a slight threat, and mouth fully open, indicating a serious threat). Cells which responded to different identities were found in Area TE, whereas those which responded to different expressions were found in the STS (as others have noted, probably in Area TPO). Area TEM contained both types of cell.

To show that a face cell responds differently to pictures of different individuals does not demonstrate its role in face recognition. Two extreme views of how individuals might be identified at the physiological level are that of the ‘grandmother’ cell, a neuron which responds only when a particular female relative appears in its receptive field, and that of distributed coding, in which the unique visual properties of an individual are encoded by the pattern of activity in a network of neurons. Imagine a neuron which encoded eye separation. This would respond differently to pictures of different individuals, if their inter-ocular separation happened to be different, but by itself the neuron would not identify those individuals. There is evidence that faces may be identified by a type of distributed coding across several neurons. Young and Yamane (1992) showed that face neurons involved in identification appear to encode combinations of features such as the distance between eyes and mouth and other features. The authors correlated the responses of different cells to a range of faces and found evidence that faces are coded by combinations of activity in different cells. They suggested that individual faces may be coded by a small number of dimensions, partly because of the high correlation between many aspects of a face, so that, for example, the larger the distance between the eyes, the larger the distance between other features is likely to be. This type of coding, in which a few
dimensions of an object are encoded by a small number of neurons is known as sparse coding, and falls somewhere between an encoding based on grandmother cells and one based on a large network of neurons.

The third line of evidence for a specific face area in the human brain comes from functional imaging, especially fMRI. The problems which arise in establishing that a particular area is selectively responding to faces include those which we have already considered with respect to face aftereffects and to the responses of single neurons to faces. Kanwisher et al. (1997) mention four such problems:
1. Is the response really to faces or to some low-level feature(s) which forms part of a face?
2. Could the response be due to visual attention which may be strongly engaged by faces?
3. Could the response be due to the structural similarity of faces, which is greater than that likely to be found in the set of control stimuli?
4. Could the response arise from the recognition of any animate (or human) object?

In their fMRI study, these authors first looked for areas in the occipital and temporal lobes which responded more strongly to photographs of faces than those of other objects. They found only one such area, which was consistently activated across all their participants, namely the right fusiform gyrus (and/or the neighbouring sulci), which they called the fusiform face area (FFA – see Figure 12.12). Because, as we shall see, the role of this area is still controversial, it will be referred to here as the right fusiform area (RFA), an anatomical rather than a functional label. Having identified the RFA in their participants, they then presented a range of stimuli, looking specifically for activity in the RFA, in an attempt to rule out several explanations.

One comparison was between two-tone faces (images containing only black and white; see Mooney faces in Figure 12.13) and scrambled versions of those faces in which the various features had been moved around. Another comparison was between grey-scale images of faces and grey-scale images of houses. The results were clear-cut: faces evoked more activity in the RFA than did either scrambled faces or houses. This seems to rule out the ideas that the RFA is

Figure 12.13 Mooney faces. This technique for representing faces was devised by Craig Mooney (1957). All luminance values in the original image below a certain value have been converted to black and all those at or above that value to white. Whereas the left-hand image above may be perceived as a face (even if not immediately), the right-hand image (an inversion of the left-hand image) is much harder to see as a face.
responding to low-level features rather than to faces, and that it is the similarity between faces which is responsible rather than the faces themselves.

The authors also compared responses to passive viewing of hands and of faces, and responses during a so-called ‘1-back’ task, in which participants had to judge whether the faces or hands in two successive presentations were the same or not. The aim of this second task was to make participants attend closely to the images. The article reports that the task was harder for hands than for faces, so that attention to hands was likely to be even greater than to faces. Again, the results were clear: faces evoked more RFA activity than did hands, even when attention to both types of stimuli was required. This suggests that the RFA is not responding more strongly to animate objects or to human body parts, and is not responding more strongly to faces simply because they engage more attention during passive viewing. Taken together, this evidence is certainly consistent with the label ‘fusiform face area’ for a region in the right fusiform gyrus.

In a further study, Kanwisher et al. (1998) studied the effects of inverting two types of facial image on RFA activity. The images were either gray-scale or Mooney faces, images in which whole areas have been transformed to either black or white. An inverted gray-scale face is still perceived as a face, even if hard to recognise, whereas an inverted Mooney face is very much harder to see as a face (see Figure 12.13). During a 1-back task, inverted grey-scale faces elicited only slightly (15%) less RFA activity than did upright faces, even though recognition performance dropped from 91 to 57%. However, inverted Mooney faces produced a much larger reduction (39%) in RFA activity in a 1-back task, compared with that from upright Mooney faces. This provides further evidence that the RFA is concerned with face perception, rather than the perception of low-level features. It also suggests that even inverted faces engage face-specific mechanisms, as suggested by the observation that inverted faces look like faces, even if they are hard to recognise.

It should come as no surprise that this interpretation of the role of the brain area centred on the right fusiform gyrus has been questioned in two ways which will be familiar from earlier sections of this chapter. First, just as for behavioural studies, as for example on the face inversion effect, it has been suggested that imaging results show that the RFA is involved in expert visual recognition of examples of a range of object types, not only faces. This is supported by fMRI work showing that the RFA is less active when naïve participants view Greebles than when they view faces, but that RFA activity evoked by Greebles rises to that evoked by faces after participants have been trained in Greeble recognition (Gauthier et al., 1999, 2000).

Second, similarly to the idea of Young and Yamane (1992) about how faces might be coded by several single neurons, it has been suggested that a range of visual stimuli, not just faces, are encoded by the distribution of activity in temporal cortex (Haxby et al., 2001), including the right fusiform area. Haxby et al. called this idea ‘object form topography’, and its central notion is that what distinguishes one set of objects from another is the relative activity in these different areas. In a test of the idea using fMRI, these authors presented
images of a range of objects: cats, houses, chairs, scissors, shoes, ‘nonsense images’ (streaks and whorls), as well as human faces. Participants performed a 1-back task (designed to promote attention to the images) in which they had to judge whether the current image was the same as or different from the previous image. In their analyses, the authors first identified voxels in which the response differed according to stimulus category, and then divided the fMRI activity in those voxels for each category into two sets, based on odd and even runs in the scanner. They could then ask whether the correlation between the odd and even runs for a particular category was higher than that between categories. If the pattern of activity was a reliable indicator of category, within-category correlations should be high and, if it was a reliable discriminator of category, between-category correlations should be low. This is just what was found. Indeed, different categories could be reliably distinguished by this method with an accuracy of 97%. However, it could be argued that the pattern of activity was dominated by activity in the brain area maximally sensitive to a category (in the case of faces, the RFA) with no reliable information being carried by regions maximally sensitive to other categories. To test this idea, the authors re-analysed the data after removing the activity in the maximally sensitive area for each category. The reliability of distinguishing categories was very similar (94%), suggesting that, for example, areas other than the RFA were contributing to face identification. Although these findings suggest that important information about the identity of objects is carried by the pattern of activity in temporal cortex, they do not reveal what this information is. However, when Haxby et al. compared the activity evoked by grey-scale images of faces, houses, and chairs with line drawings of those objects, they found similar patterns of activity evoked by the different types of representation, which again could be used to distinguish between categories with 96% accuracy. This suggests that the pattern of activity does not represent low-level features of the image (though it still leaves open the important question of how objects are represented in temporal cortex).

KEY POINTS

- Neuropsychological studies suggest that mechanisms within the temporal lobe are important in face recognition
- Some single neurons in monkey superior temporal sulcus respond more vigorously to faces than to other objects (by a factor of between 2 and 10)
- An area in the right fusiform gyrus responds more vigorously to faces than to other objects, but also more vigorously to Greebles, after training
- There is evidence from fMRI that a network of areas underlies recognition of face and of other objects. Different types of object evoke different distributions of activity within this network
CONTROVERSIES IN STUDIES OF FACE RECOGNITION

Face recognition is a vibrant area of research, and many of the conclusions and some of the findings are controversial. For example, Robbins and McKone (2007) questioned whether the (undoubted) changes in recognition performance with expertise really reflect the development of configural processing. They failed to replicate the finding of Diamond and Carey (1986) that dog experts showed a larger inversion effect than novices in dog recognition (and point out that, in any case, effects of expertise on recognition could reflect improved matching of individual features). Robbins and McKone also found no Composite Dog Effect in dog experts (though the Composite Face Effect has been taken as strong evidence for configural processing). They argue that expertise in object recognition does not result in ‘face-like’ processing of non-face objects. Their conclusion is disputed by Gauthier and Bukach (2007), who suggest, for example, that it is based on selective attention to behavioural data, to the exclusion of work on brain imaging, and that Robbins and McKone’s measure of the Composite Face/Dog Effect may not be optimal. McKone and Robbins (2007) vigorously defended their original conclusion.

It may be that some of the controversies are more apparent than real. Ishai (2008) argued in support of a distributed network of areas underlying face processing, in which ‘core’ regions (in the RFA, for example) interact with ‘extended’ regions in prefrontal cortex or the limbic system in a manner which depends on the context. One line of objection to this idea (Wiggett and Downing, 2008) is that some of the brain regions in the proposed network respond also to non-face stimuli, and might, for example, be processing the emotion expressed by the face. They argue that to describe such regions as ‘face-processing’ is to over-extend the idea of a network. However, the disagreements between these researchers may not be fundamental, and to some extent reflect differences of emphasis. In contrast, Pitcher et al. (2009) concluded that their study created difficulties for the idea of a distributed network of regions involved in recognition of different types of objects. They used fMRI to identify in each participant the right occipital face area (rOFA – which responds well to faces), the right extra-striate body area (rEBA – which responds well to human bodies), and the right lateral occipital area (rLO – which responds well to other objects) (see Figure 12.12 for the location of these areas in the human brain). Although this was not demonstrated in their study, there is evidence from other fMRI studies that, as well as responding to their preferred stimuli, each region responds, though less well, to the other classes of image. The participants then performed a series of discrimination tasks while TMS (transcranial magnetic stimulation) was applied to one of these areas. They found that TMS had selective effects: over rOFA, it affected discriminations involving faces, but not bodies or objects; over rEBA, it affected discriminations involving bodies, but not faces or objects; and over rLO, it affected discriminations involving objects but not faces or bodies. The authors argued that, if these regions formed part of a network in which objects were encoded by the distribution of activity, discrimination of all three types of image should have been affected by TMS at any of the sites, and that their data supported the idea of individual modules,
each specialised for processing a particular type of image. Although this study casts doubt on the notion of a widespread network of areas encoding faces (and other objects), it may still be the case that (say, within the RFA) there is a network of neurons which encode faces by representing them along a small number of dimensions (and possibly overlaps with networks encoding other objects). There is evidence from high resolution fMRI that the RFA is not uniform: it certainly contains many sub-areas which respond to faces, but interspersed with these are sub-areas which respond to other objects (Grill-Spector et al., 2006). Improvements in imaging technology are likely to produce greater understanding of the processing of faces and other objects within the RFA.

**KEY POINTS**

- It is controversial whether developing expertise in recognising non-face objects includes the development of configural processing
- There is evidence against a widespread network for recognition of different objects, but this does not rule out a local network within the right fusiform gyrus for the recognition of various types of object (including faces)

**OVERVIEW OF STUDIES OF FACE RECOGNITION**

Despite areas of disagreement, some aspects of face recognition are well understood. The internal features of a face seem particularly important in recognition, and these are processed as a configuration. Individual faces seem to be coded as deviations from a norm, based on previously seen faces, and biased towards recently seen faces. Studies of single cells in monkeys suggest that individual faces might be represented in the brain by differences from the norm along a small number of dimensions, such as distance between the eyes. There is at least one site in the brain which is specialised for face recognition, damage to which impairs face recognition.

**SUMMARY**

1. Although recognising that something is a face appears to be similar to the recognition of other objects, identifying an individual face appears to be different, and to be based on configural processing (as shown by the Composite Face Effect).
Inverting the images impairs recognition of faces more than of other objects, unless the observer is an expert in recognising that class of stimuli.

Face aftereffects suggest that individual faces may be remembered as differences from a norm, or average, face, based on faces previously seen.

Face aftereffects reflect changes in face-processing mechanisms, not in their low-level inputs.

There is at least one site in the brain which is heavily involved in face processing, can be identified by fMRI, and damage to which impairs face recognition (prosopagnosia).

The differences from the norm by which a face is encoded may be relatively few (and include, for example, the distance between the eyes).

TEST YOURSELF QUESTIONS

1. You should now be able to say in your own words what each of the following terms means:

   - Primal sketch
   - Recognition by Components
   - Geons
   - Inversion effect
   - Configural processing
   - Greebles
   - Face space
   - Anti-face
   - Morphing
   - Face aftereffects
   - Prospagnosia
   - Object form topography

2. What is the Composite Face Effect? Why is it important for theories of face recognition?

3. To what extent do face aftereffects support the idea of norm-based coding of faces?

4. Is there a brain region specialised for face recognition?

SUGGESTIONS FOR FURTHER READING

Biederman I, Kalocsai P (1997) Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London B* 352: 1203–1219. Summarises evidence that face identification may be different from recognising that an object is a car (or, indeed, a face).