

CHAPTER 4

RECOGNITION

Graham Pike and Nicola Brace

1 INTRODUCTION

In the last chapter on perception, we explored some of the cognitive processes involved in forming a mental description of the environment based on input from the senses. As well as being able to determine the position and shape of the objects around us, it is also possible to recognize *what* we are seeing. Unless we fully accept Gibson's concept of affordance (and it's safe to say that we don't), there must be another step: another set of processes that transform the basic descriptions of objects generated by analysing the retinal image into objects that are familiar to us and that we can recognize as specific objects, such as a mobile phone, coat, or cow.

The same is, of course, true of our other senses; for example, when we listen we may hear music, car engines, and voices. Again, there must be cognitive processes that somehow transform the auditory input of sound waves into what we recognize as an environment of voices, music, and cars.

Let's stop for a moment and consider the basic steps that might be involved in the process of visually recognizing an object:

- First, there must be processes that are able to construct an internal representation (referred to as

a 'description') of the object, based on the information in the retinal image.

- Second, there must be processes that are able to store this description so that we can recognize the object if we see it again.
- Third, there must be processes that somehow compare the description of the object that we can currently see to the descriptions of objects that we have stored.
- Lastly, it is very likely that we have seen objects from many different angles, yet are able to recognize them regardless of the current angle of view. As we shall see, the nature of the mechanism that allows us to do this is an important and much debated point.

A basic diagram displaying the recognition process is provided in Figure 4.1. In one sense, the process of recognition is the process of generating and comparing descriptions of objects that are currently in view with descriptions of objects that we have seen previously. It is worth noting that this is a very simplistic way of viewing and describing recognition, and in Section 2 we shall look at some of the problems with this simplistic approach.

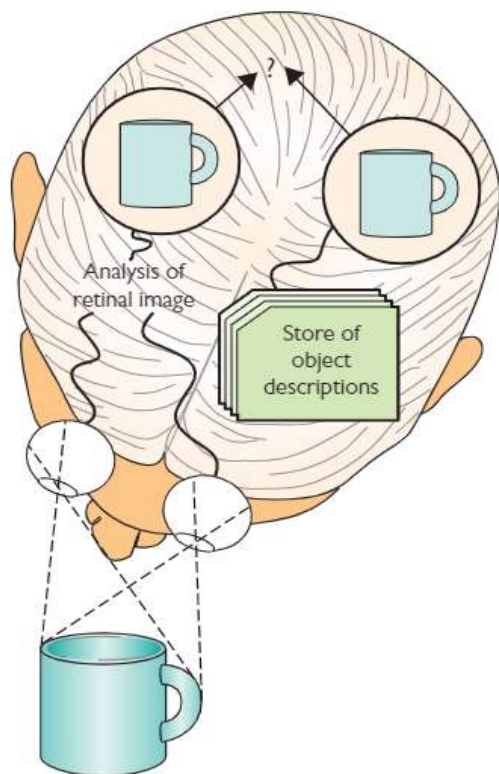


FIGURE 4.1 The basics of the recognition process.

1.1 Recognition in the wider context of cognition

In Figure 4.2, we can see how Humphreys and Bruce (1989) summarized the way in which object recognition fits into a wider context of cognition that includes perception (perceptual classification), categorization (semantic classification), and naming. As you can see from Figure 4.2, the first stage in the process is the early visual processing of the retinal image. One example of this form of processing is that which produces Marr's full primal sketch (Marr, 1982). In the second stage a description of the object is generated,

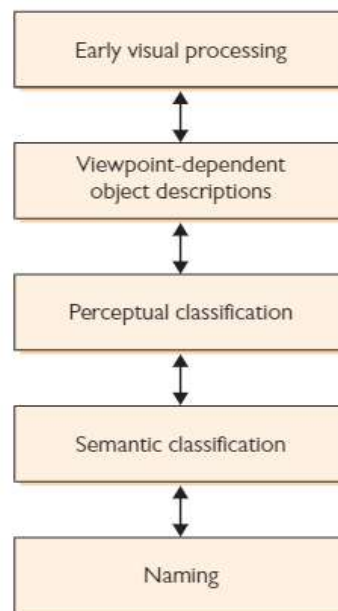


FIGURE 4.2 Model of object recognition suggested by Humphreys and Bruce (1989).

but this description is dependent on the viewpoint of the observer. This stage is therefore similar to what Marr (1982) referred to as the $2\frac{1}{2}$ D sketch.

Humphreys and Bruce refer to the next stage as 'perceptual classification' and it is really this stage that we have been discussing so far in this chapter. Perceptual classification involves a comparison of the information regarding the object in view with descriptions of objects that have been stored previously. It is at this stage that the object is 'recognized'.

Once the object has been recognized, or perceptually classified, it can then be 'semantically classified'. This process, also referred to as 'categorization', is examined in the next chapter. Once this stage has been achieved, the object can then be named, aspects of which will be examined in the later chapters on language.

SUMMARY OF SECTION 1

- As well as being able to determine the location and shape of an object, or the location and pitch of a sound, we also have to be able to recognize what they are.
- A basic model of recognition requires that a description from sensory input is generated and compared with descriptions stored in memory.
- Recognition must come after the initial processes of perception and before the stages in which an object can be first semantically classified and then named.

2 DIFFERENT TYPES OF RECOGNITION

As we have stated above, the view that recognition involves comparing an object description generated from the retinal image to descriptions stored in long-term memory is very simplistic. In fact there are quite different *types* of recognition, depending on what it is we are trying to recognize and how we go about trying to recognize it. Throughout this chapter we shall be exploring these different types of recognition and examining some of the issues that suggest the process of recognition is far more complex than the simplistic model presented in Figure 4.1 suggests.

2.1 Object and face recognition

The end point of Humphreys and Bruce's (1989) model of recognition (Figure 4.2) is the naming stage. Naming, of course, is not a necessary component of being able to recognize an object: even if an animal has no capacity for language, it can still recognize objects. But the names we give things do provide a clue to the fact that there are different types of recognition.

Try Activity 4.1. In completing this, you may have realized that the name you provided for the left image was the category to which the object belonged, whilst the name for the right image corresponded to an individual rather than a category (i.e. you did not name the image 'a face').

Naming reveals that it is possible to recognize objects in different ways. When we see objects such as fruit and furniture we tend to concentrate on which category they belong to, and when we provide names for them, these are usually the name for that category. Thus, we are making **between-category** distinctions such as 'that object is an orange and that one is a table'. However, when we see a face we often do more than

recognize that the object belongs to the category of objects known as 'faces'; we also work out *whose* face it is. In other words, we make a **within-category** distinction.

The difference between within- and between-category recognition is one reason why **face recognition** is generally researched as a separate topic from **object recognition**. In addition, there are some issues that are unique to face recognition, such as:

- The internal features of a face can move, which changes the appearance of the face.
- This movement can serve to express emotional and social cues.
- Faces can change quite dramatically over time, due to ageing or haircuts for instance.

As well as distinguishing between face recognition and more general object recognition, it is possible to identify a number of different types of face recognition. One such distinction is between recognizing familiar and unfamiliar faces. Pike *et al.* (2000) reported that people were often able to identify E-FIT see images even when other participants had rated them as a poor likeness. However, like the E-FIT you will see in Activity 4.2, these were images of famous people whose faces would have been familiar to the participants. Considerable evidence suggests we are not so accurate at recognizing even real faces that are not so familiar to us. For example, many witnesses express uncertainty when asked to identify the perpetrator of a crime from a line-up (Pike *et al.*, 2001). Even when the anxiety of the witness is reduced by using a video identification parade, identification accuracy is far from perfect (Kemp *et al.*, 2001).

A second distinction that applies to types of face recognition is that between recognizing whose face you are looking at and recognizing what emotion it may be portraying. You can imagine that the importance of

* ACTIVITY 4.1

Figure 4.3 shows two images. See if you can name them.

You probably provided the names 'apple' and 'Sigmund Freud'. These are evidently two different

types of name, but can you describe why these two types of name are so different? (Hint: think about how many different apples and Sigmund Freuds there are.)

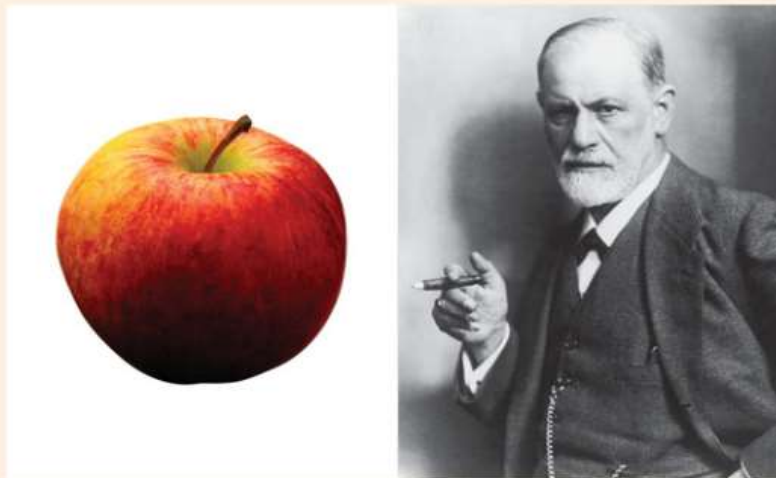


FIGURE 4.3

* ACTIVITY 4.2



FIGURE 4.4

Can you identify the person depicted in the three images shown in Figure 4.4?

COMMENT

The images are of Paul McCartney and you were (probably) able to recognize him from all three images, even though there are some quite obvious

differences in appearance. In fact, you were probably able to recognize the E-FIT image of him (right-hand image), even though this is constructed by combining together features from several other faces. So, we can recognize a face that is familiar to us even when quite large changes have been introduced.

faces in conveying emotional state and in facilitating social interactions has led us to develop some very sophisticated cognitive processes for interpreting facial expressions. In fact, we are able to judge the emotion being displayed on a face with great accuracy (the cognition involved in perceiving emotion is considered in Chapter 19) and are very sensitive to eye movements in those around us. It is tempting to think that we may have evolved a specific set of cognitive processes for recognizing faces and the emotions they express because of the social importance of this information. However, there is evidence (Young *et al.*, 1993) that although we do have specific processes for recognizing emotions, these processes are not involved in recognizing identity. We shall return to the difference between emotion and identity recognition later in this chapter, but logically you can see that you need to be able to tell whether someone is angry or happy regardless of whether you can recognize them or not. Likewise, you need to be able to recognize who someone is regardless of whether they appear happy or angry.

The question of whether faces are recognized by the same cognitive processes that are used to recognize other objects has been at the centre of a great deal of research. Although a definitive answer as to just how different face recognition is from general object recognition has yet to be provided, the two have tended to be treated as different areas of research. Because of this, we have divided this chapter into two main areas of discussion. The first (the rest of Section 2 and Section 3) will look at theories of how we recognize objects, and the second (Sections 4 to 7) will look at models of face recognition and examine in more depth the question of whether faces are recognized by special processes.

2.2 Active processing – recognizing objects by touch

One limitation of the basic recognition procedure we suggested in Section 1 is that it treats recognition as a passive process. Gibson (1986) stressed that perception is an active process and that we are beings who

interact with and investigate the environment. In examining how Gibson's idea of active perception might apply to recognition, we will temporarily switch modalities from vision to touch. One reason for concentrating on touch is that purely passive object recognition through touch would be almost impossible. Although there may be some objects that you can recognize if they were simply placed on your hand, most objects would require exploration. We have evolved sophisticated processes for exploring the environment and objects using touch in very exact and careful ways.

First, we have tremendous control over our hands, so that we can both move our fingers precisely and also apply varying degrees of pressure to objects in a very measured way. This is done by employing a **feedback** system, whereby information from touch receptors allows the brain to control the location and amount of pressure applied by the fingers. As well as being able to regulate touch precisely, we can also pinpoint the location of our limbs with great accuracy via receptors inside our muscles and joints. This information about limb location is known as **kinesthesia**, and it can be combined with information from the touch receptors to guide our hands and fingers. Of particular importance are the relative positions of your fingers as they touch the object, their orientation to your hand, and the position of your hand in relation to your arm and of your arm in relation to your body. The processes that allow us to keep track of the relative locations of all our limbs are known as **proprioception**.

So, at every moment that we are touching an object, we know the exact position of our fingers (kinesthesia) and what the object feels like at that point (touch receptor information). The information gained from this combination is referred to as **haptic information** and it can be used to generate a description of an object.

Lederman and Klatzky (1987) found that there was considerable consistency in the way in which people used their hands in order to gather haptic information. They described how participants tended to use a series of exploratory procedures when investigating an object with their hands. Lederman and Klatzky (1990) went on to study these exploratory procedures in more depth and described how each particular

* ACTIVITY 4.3

Ask someone to place a variety of objects within easy reach of you (you can do this yourself if you wish). Ask them to choose objects that differ in shape, texture, and weight. Close your eyes and pick up each object in turn and try to work out what it is. As you do this, try to make a mental note of the different movements that your hands

make and what each movement tells you about the object.

Table 4.1 gives a list of some of the hand movements reported by Lederman and Klatzky (1987), along with the information that these exploratory procedures tend to reveal. Did you find yourself using these movements?

TABLE 4.1 The information revealed by exploratory hand movements

Movement	Information
Enclose object in hand(s)	Overall shape
Following contours with fingers	More exact shape
Lateral motion with fingers	Texture
Press with fingers	Hardness
Static contact with fingers	Temperature
Unsupported holding	Weight

Source: based on Lederman and Klatzky, 1987, Table 1, p.345

procedure could be used to derive a certain type of information that was useful for recognizing an object. For example, if shape was important in recognizing the object people tended to move their fingers around the object's contours, and if texture was important they would move their fingers across the surface of the object.

Although haptic perception can be used to recognize objects, visual recognition has the obvious advantage that it can be used for distant objects that are out of reach. It also tends to be far quicker and more accurate in processing information about shape, particularly complex 3D shape (Lederman *et al.*, 1993). But visual perception is not so useful when it comes to judging the weight and texture of an object.

So, haptic perception is a very useful source of information and can be used to recognize certain objects. The study of haptic perception also serves to demonstrate that recognition is not necessarily passive and that much can be gained from considering it

as an active process. Nor is active perception limited to touch. You saw in the last chapter how your interpretation of the impossible triangle (Figure 3.24) kept changing as you visually explored the object, corner by corner.

2.3 Recognizing two-dimensional objects

Another way of distinguishing between types of recognition is according to whether the object in question is three-dimensional (3D), such as the book in front of you, or two-dimensional (2D), such as the words in front of you. The difference between 2D and 3D object recognition takes on added significance when you consider that the description generated from the retinal image will in essence be 2D, whilst most objects tend to be 3D. In fact, much of the early research conducted on recognition processes was focused on how simple, two-dimensional 'patterns' are recognized. Although it can be argued that this work tells us little about how complex, three-dimensional objects are recognized, it does serve to highlight some of the problems that are inherent in any approach to object recognition.

By far the simplest model of visual pattern recognition postulates **template matching**. This is the idea that we have a large number of templates stored in long-term memory against which we compare the patterns we come across. For example, a template would exist for every number from 0 to 9 and for every letter from



FIGURE 4.5 Different alphanumeric characters that share similar properties.

A to Z. The problem with this theory is that it cannot cope with the enormous variation in the actual patterns that are used to represent even simple things such as alphanumeric characters. For example, in Figure 4.5 the top row contains examples of the letter ‘R’ and the bottom row contains examples of letters that each share many similar properties with the specific example of an ‘R’ immediately above it. Although we do not have any great difficulty in reading these letters, it is hard to see how a simple template could be created that would accept every example in the top row as a letter ‘R’ and reject every example in the bottom row.

If the problem with template matching is that the template cannot deal with variation in the stimulus it has to recognize, then we have to look at some way of representing objects that is not so reliant on the exact visible pattern. One way of doing this is to try to extract the key characteristics or features of an object. In the case of alphanumeric characters, these features could be the number of curved and straight lines and the relationship between them. An ‘O’ might therefore be represented as a single continuous curve, a ‘P’ as one vertical line and one discontinuous curve, and a ‘T’ as one horizontal and one vertical line that form two right angles.

One of the most influential **feature recognition theories** is the Pandemonium system, so called because processing units known as ‘demons’ were used to detect each feature. This system was designed as the basis for a computer program to decode Morse code signals (Selfridge, 1959) and was later adapted by Neisser (1967) to recognize alphanumeric characters. Although Pandemonium systems have been useful in recognizing simple, highly constrained patterns, they do not pro-

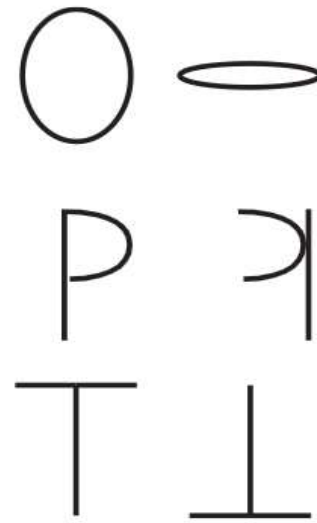


FIGURE 4.6 Examples of different patterns described by the same key features.

vide a particularly useful model of human object recognition. A central flaw in feature recognition theories is that describing an object in terms of a list of key features does not capture the structural relations *between* features. If you look back at the feature-based descriptions provided for an ‘O’, ‘P’, and ‘T’ above, you will see that these three descriptions could also apply to the figures presented to the right of each letter in Figure 4.6, meaning that these shapes would be misidentified as letters.

An approach that has had more success in explaining how both simple patterns and more complex objects might be recognized is that based on **structural descriptions**. Structural descriptions are made up of a series of propositions, based both on a description of the elements that comprise the object and the structural relations between them. Thus, the structural description of a letter ‘L’ might contain the following propositions:

- There are two lines.
- There is one horizontal line.
- There is one vertical line.
- The horizontal line supports the vertical line.
- The horizontal line extends to the right of the vertical line.
- The horizontal and vertical lines are joined at a right angle.

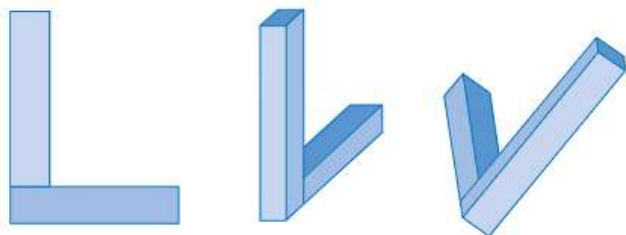


FIGURE 4.7 Three representations of a 3D 'L' shape.

Although the propositions stated above are expressed in language, they can be equally well expressed in other forms of symbolic representation, such as that used in a computer program.

One key advantage that structural descriptions have is that it is possible to see how they could be applied to three-dimensional objects. Consider the three representations of the character 'L' in Figure 4.7. Both template matching and feature recognition theories would recognize the representation to the left as being an 'L', but would immediately reject the other two. However, the two forms of the letter 'L' on the right of Figure 4.7 do share a similar structural description once we consider their three-dimensional properties.

However, in order to obtain a description that includes elements of three-dimensional structure, we must be able to turn the 2D retinal image, which is dependent on the particular view that the observer has of the object,

into a 3D description that is centred not on the viewer but on the object itself. This, as you might expect, requires an even more sophisticated means of describing objects, and is the focus of the second half of Marr's theory of vision – which we shall look at in Section 3.

2.4 Object-centred vs viewer-centred descriptions

One of the most fundamental problems in recognizing an object is that it is possible to view an object from many angles. As we have seen, any theory that treats an object as a simple pattern is likely to fail when applied to a 3D object (as with the 'L' in Figure 4.7). Consider writing a very simple computer program based on recognizing an object by matching patterns. As an example, Figure 4.8 contains a conceptual diagram of how a computer might be programmed to recognize a coffee mug.

But coffee cups are actually 3D objects and can be viewed from many angles. Let's see how our simple computer program would cope if we turned our coffee cup so it was facing the other way. As you can see from Figure 4.9, the program has decided that, as the patterns do not match, the object is NOT a coffee cup.

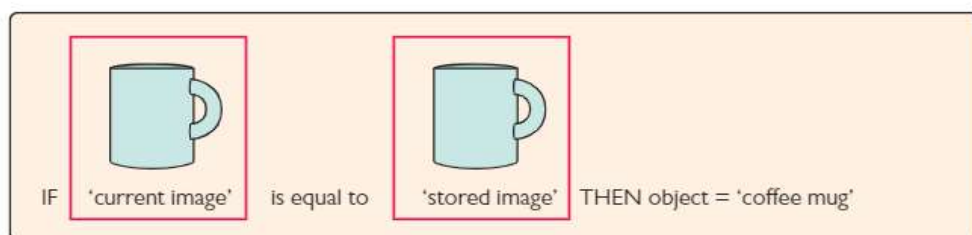


FIGURE 4.8 A simple program for recognizing an object.

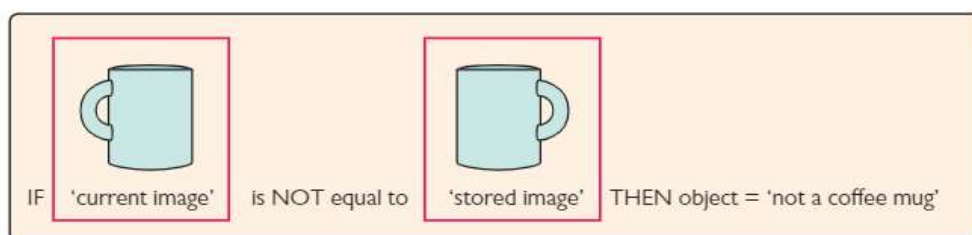


FIGURE 4.9 A simple program failing to recognize an object from a different viewpoint.

The failure of the simple program to deal with a small change in viewpoint is obviously an unacceptable flaw in any system that wishes to interact with its environment. Instead of being reliant on seeing objects from just a single viewpoint, the process of object recognition must somehow be based on descriptions of objects that allow recognition to take place *independent* of viewpoint. In fact, these processes must be tolerant of any naturally occurring change, not just changes in viewpoint. This is a very

important point and one that is central to the study of object recognition.

Marr (1982) conceptualized the problem of viewpoint as that of turning the **viewer-centred description** of the object that was formed in the 2½D sketch (see Chapter 3, Section 4.3) into a **3D object-centred description** that would allow the object to be recognized despite changes in viewpoint. In the next section we shall look at how Marr suggested this might happen.

SUMMARY OF SECTION 2

- There are different types of recognition, which depend on what is being recognized and how.
- Object recognition tends to be based on making between-category distinctions and face recognition on making within-category distinctions.
- Face recognition tends to be researched apart from more general object recognition because faces can convey social and emotional information and their appearance can change.
- Recognition is not entirely a 'passive' process and can involve an active exploration of the environment. This is particularly true of haptic recognition, in which objects are recognized by touch.
- One key problem facing any theory of visual recognition is that the retinal image is essentially 2D, but objects are 3D.
- Early theories that concentrated on recognizing 2D patterns, such as template matching and feature recognition theories, are therefore not particularly useful models of human recognition.
- Theories based on abstracting a structural description of an object are better able to cope with 3D objects.
- As a 3D object can be viewed from many angles, our recognition system must be able to turn an object description centred on the viewer into one centred on the object.

3 RECOGNIZING THREE-DIMENSIONAL OBJECTS

As we saw in the previous chapter, in the first part of Marr's theory of perception, early visual processing of the retinal image eventually leads to the generation of the 2½D sketch. But the surfaces and objects in the 2½D sketch are described in relation to the viewpoint of the observer and are therefore viewer-centred descriptions. As we saw in the pre-

vious section, viewer-centred descriptions are of little use in recognizing real objects that can be seen from any angle and any distance. The second half of Marr's theory therefore concentrated on recognising objects from any angle by working out how to turn the 2½D sketch into a 3D object-centred description.

If it were not possible to generate a 3D object-centred description, the only way of accurately recognizing objects would be to store a very large number of viewer-centred descriptions. Although there are theories that have taken this approach, for now we will concentrate on the idea that recognition is best subserved by a single representation of an object that can be used to recognize it from any angle.

Marr and Nishihara (1978) suggested that objects could be represented by generating a 3D object-centred description that would allow the object to be recognized from virtually any angle. They proposed that this description was based on a **canonical coordinate frame**. This basically means that each object would be represented within a framework that was about the same shape as the object. You could imagine the representation of a carrot as being a cylinder that tapered towards one end.

This procedure appears at first glance to be somewhat paradoxical, as it would be necessary to know the approximate shape of the object before you could begin to recognize it! However, remember that the formation of the 3D object-centred description occurs after considerable analysis of the retinal image has already taken place, so some information as to the shape/outline of the object will already exist.

3.1 Marr and Nishihara's theory

Marr and Nishihara saw the first step in establishing a canonical coordinate frame as defining a central axis for the object in question. This is relatively easy to do

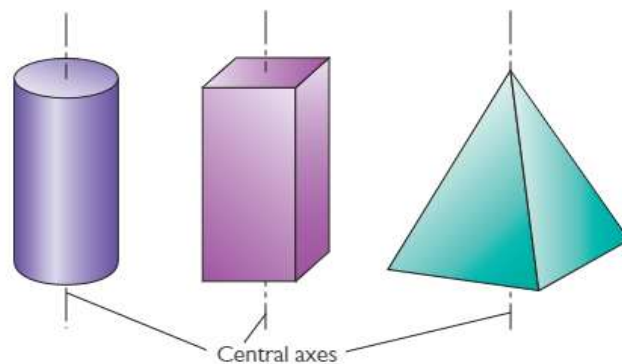


FIGURE 4.10 Locating the central axis of an object.

if the object in question either has a natural line of symmetry or has a length that is noticeably greater than its width and depth (see Figure 4.10).

In fact, the generation of the central axis is so important in Marr and Nishihara's theory that it is restricted to specific objects that can be easily described by one or more **generalized cones**. A generalized cone is any 3D shape that has a cross-section of a consistent shape throughout its length. The cross-section can vary in size but not in shape. All of the objects shown in Figure 4.11 are examples of generalized cones. Although restricting the theory to generalized cones is undoubtedly one weakness of Marr and Nishihara's theory, the basic shape of many natural objects, particularly those that grow (such as animals and plants), can be described, albeit rather loosely, in this way.

To locate the central axis of an object, it is first necessary to make use of the information contained within the $2\frac{1}{2}$ D sketch in order to work out what shape the object has. Marr (1977) suggested that it is possible to work out the shape of an object based on the object's **occluding contours** (these are basically the object's silhouette). The points on the object's surface

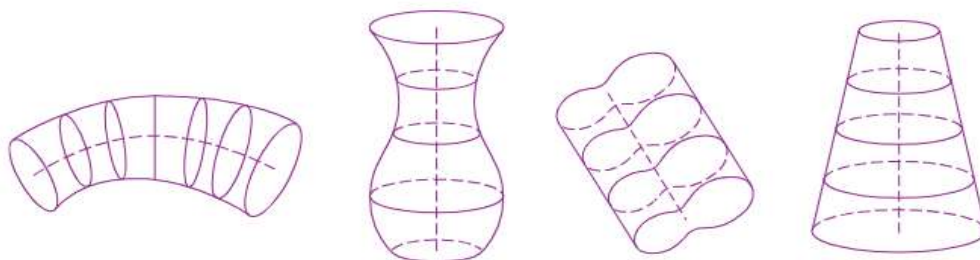


FIGURE 4.11 Four generalized cones. Source: Marr, 1982, Figure 3.59, p.224



FIGURE 4.12 *Rites of Spring* by Picasso.

that correspond to the boundary of its silhouette are of particular importance in Marr's theory, and he referred to them as the *contour generator* – because they can be used to generate the contour of the object.

As Marr (1982) points out, we seem to have no problems in deriving 3D shapes from silhouettes such as those used in Picasso's *Rites of Spring* (see Figure 4.12). However, as the silhouette of an object is two-dimensional, it is possible that it could be caused by more than one 3D object. Consider the circular silhouette (a) in Figure 4.13. This could be caused by any of the 3D objects below it (if they were sufficiently rotated), yet we tend to interpret the silhouette as being produced by the sphere (b).

Marr suggested that the problem of how we can derive 3D shapes from 2D silhouettes is solved by the visual system making certain assumptions about what it is seeing. As Marr himself said, 'Somewhere buried in the perceptual machinery that can interpret silhouettes as three-dimensional shapes, there must lie some source of additional information that constrains us to see the silhouettes as we do' (Marr, 1982, p.219). Marr conceptualized this 'additional information' as coming in the form of three basic assumptions built into the computational processes:

- Each point on the contour generator corresponds to a different point on the object.

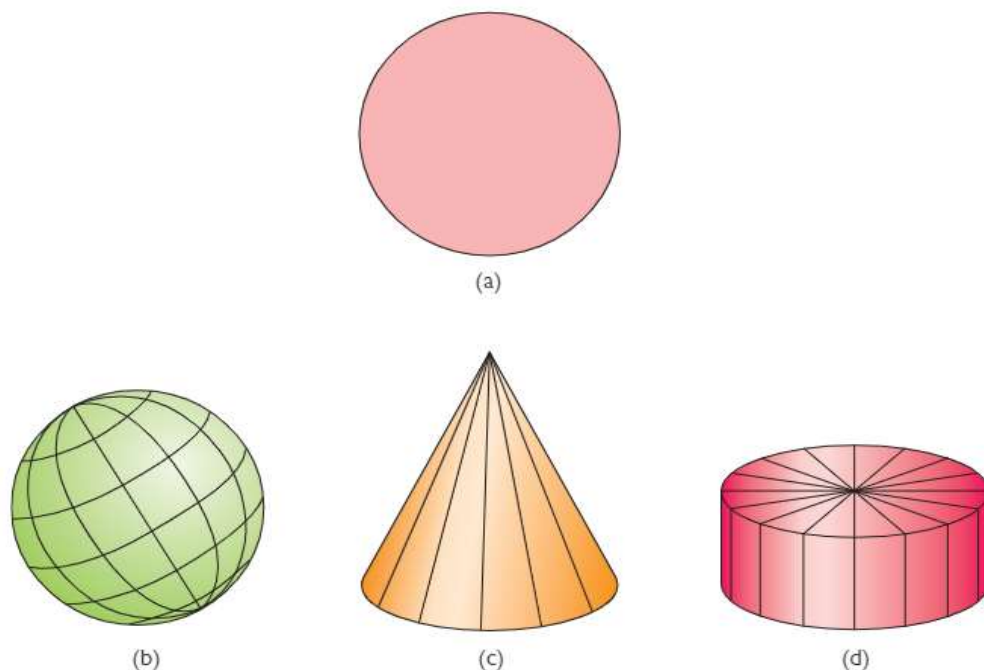


FIGURE 4.13 A silhouette (a) and three objects that could cause it (b, c, and d).

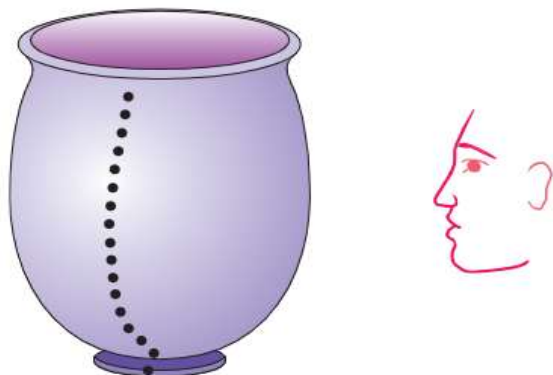


FIGURE 4.14 The black dots indicate points that lie in the same plane with respect to the viewer. Source: Marr, 1982, Figure 3.57(d), p.220

- Any two points that are close together on the contour in an image are also close together on the contour generator of the object.
- All the points on the contour generator lie in a single plane (i.e. are planar).

The first two points are relatively straightforward and the third assumption has been illustrated in Figure 4.14.

The third assumption, that all of the points on the contour generator are planar, is a vital component in Marr's theory, but it can be problematic. As we have seen, it is possible for two quite different objects to share the same silhouette and for the points on the silhouette to vary in their distance from the observer. We tend to interpret the contour on the left in Figure 4.15 as being a hexagon. However, this contour will be produced by the cube to the right. The problem is that the assumption that all the points on the contour are planar is violated by this view of the cube, as point (A) is further away than point (B). As the points on the cube's occluding contour are not planar, we tend to interpret its silhouette incorrectly.

Once the shape of the object has been derived using its contour generator, the next step is to locate the axis/axes necessary to represent the object. It is fairly straightforward to do this when the shape is simple, as symmetry usually tells us where its axis is located, but what about more complex shapes? The answer is that we often need to represent the shape using several axes, so that the object is divided into components and

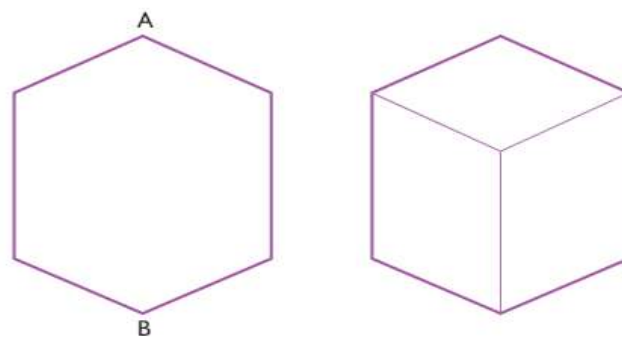


FIGURE 4.15 The contour of a cube may not be planar. Source: based on Marr, 1982, Figure 3.58, p.221

one axis is used for each component (these are referred to as component axes).

In Figure 4.16, one method of locating axes suggested by Marr and Nishihara (1978) is illustrated. The object in question is a toy donkey (a). The first step (b) involves working out areas of concavity (these correspond to parts of the contour that include a bend inwards and are represented in the figure by a '-') and convexity (parts of the contour that include a bend outwards, represented by a '+'). The shape can then be divided into sections by finding areas of sharp concavity (c) and using these to divide the object into smaller parts (d). Once the shape has been divided in this way, it is possible to represent each section via a component axis (e). These component axes can then be represented in relation to the horizontal axis of the body (f).

Figure 4.17 illustrates how it is possible to represent a quite complex object using several components, or **primitives** as Marr called them. The description of the object must allow recognition at a global level, such as being able to tell that an object is a human body, and must also incorporate more detailed information, such as the fact that a human hand has five fingers. It is therefore necessary for there to be a hierarchy of 3D models, in which each subsequent level contains a more detailed description of a specific part of the object. This means that fewer primitives will be used to represent each part of the object at the higher levels of the hierarchy.

For example, consider the description of the human body provided in Figure 4.17. At the highest level, the entire human body is described in relation to a single

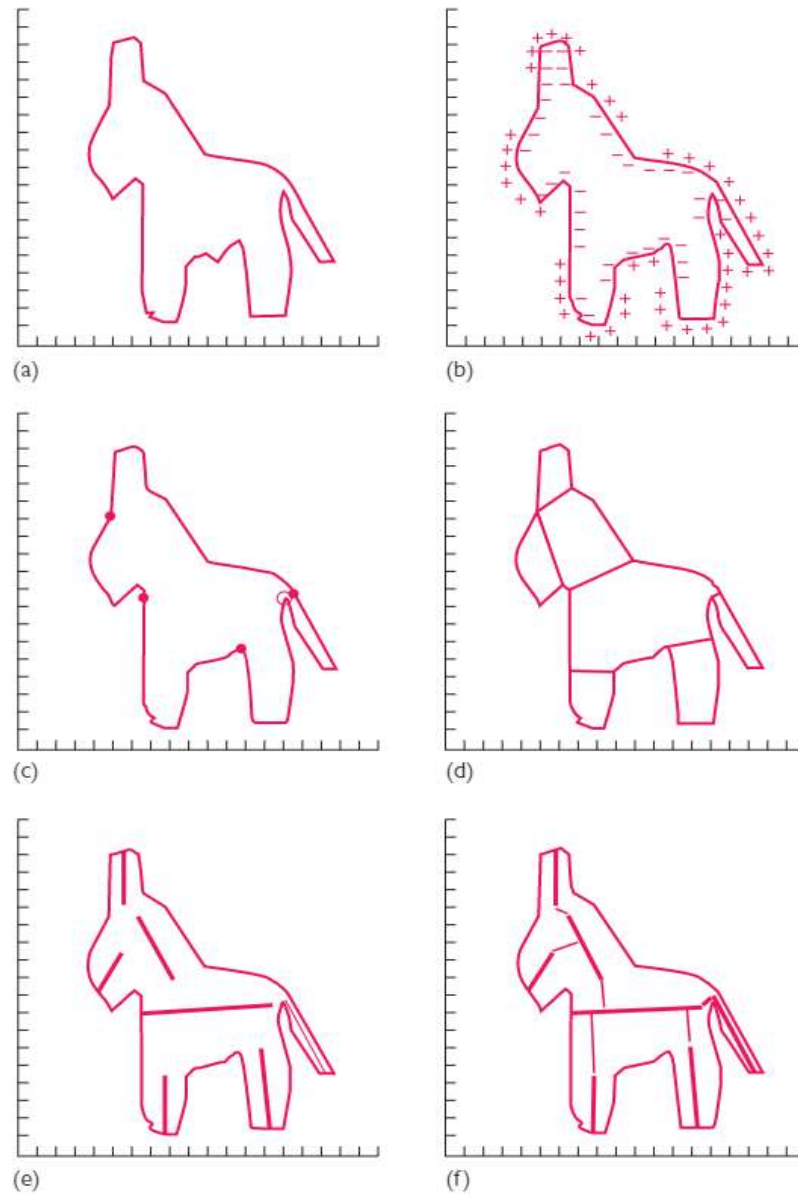


FIGURE 4.16 Locating the component axes of an object. Source: Marr and Nishihara, 1978, Figure 6

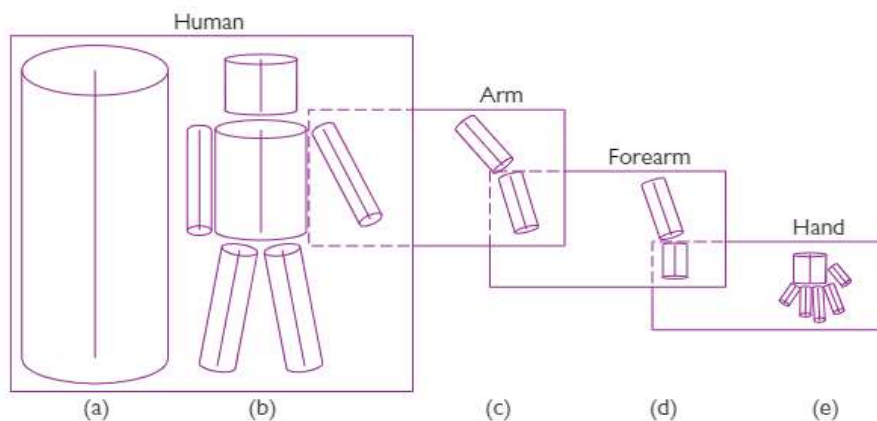


FIGURE 4.17 Marr and Nishihara's hierarchical model of a human body. Source: Marr and Nishihara, 1978, Figure 3

axis that runs through the centre of the body (a). This 3D model also contains the relative length and orientation of the axes that describe the head, torso, arms, and legs (b). However, no details regarding smaller parts (such as the fingers) are provided.

The axis that corresponds to each limb (b) is then used as the major axis for a more detailed description of that limb (c). For example, the axis of the cylinder representing the right arm is then used as the major axis to represent the upper and lower parts of that arm (c). The axis of the cylinder used to describe the lower part of the arm (c) is used as the major axis to describe

the forearm and hand (d). Finally, the axis of the cylinder used to describe the hand (d) is used as the major axis in order to describe the five fingers (e). Thus we have a 3D model description that can be used to recognize an entire human body, as well as any of its parts.

Having derived a 3D description of the object, Marr and Nishihara (1978) saw the next step in the process of recognition as comparing this to a catalogue of 3D models, formed from the 3D descriptions of all previously seen objects. The catalogue is organized hierarchically according to the amount of detail present in the model (see Figure 4.18). At the highest level the

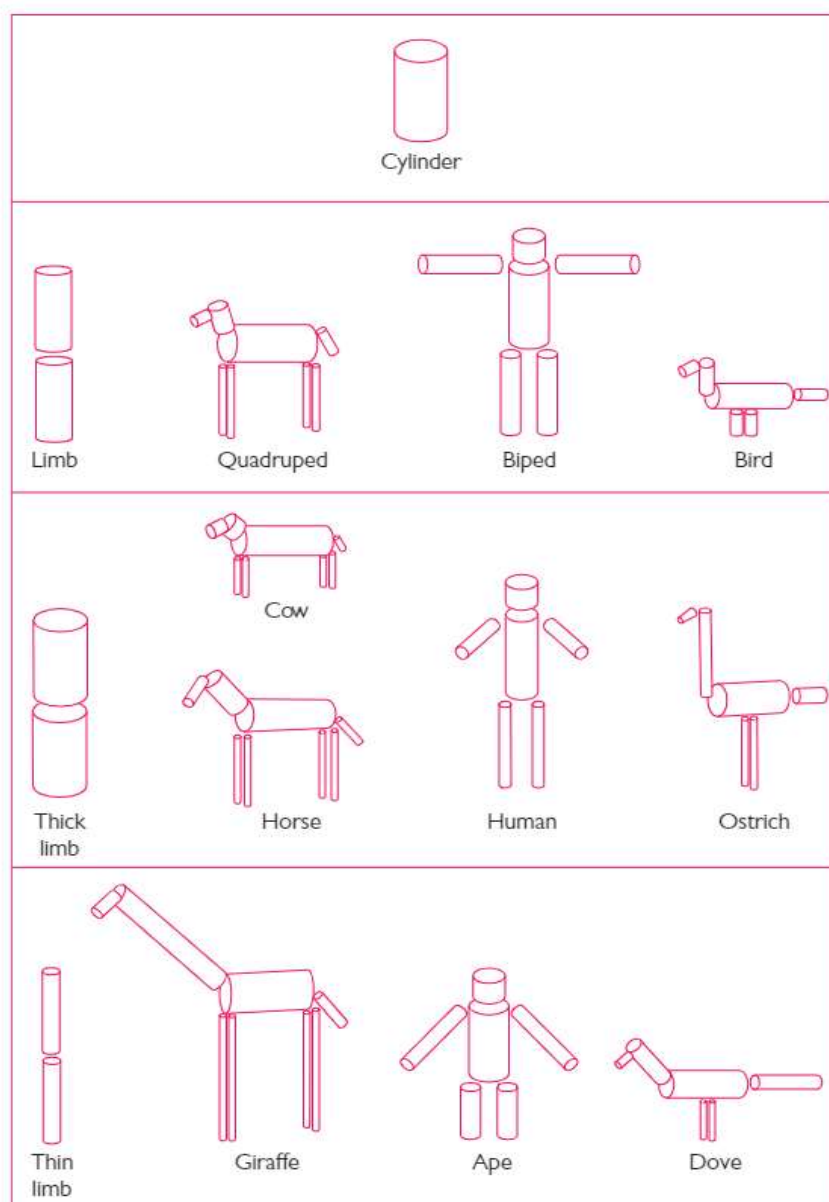


FIGURE 4.18 3D model catalogue. Source: Marr and Nishihara, 1978, Figure 8

catalogue consists of descriptions devoid of any decomposition into components. The next level contains more detail, corresponding to the number and basic layout of limbs as in Figure 4.17. At the next level even more detail is contained, such as that relating to the angles and lengths of component axes.

The 3D model generated of a new object (the target) is related to the catalogue, starting at the highest level. The target is compared to the stored models and the example it best matches is used as the basis for the next level of detail. The process stops when a level is reached that corresponds to the level of detail present in the target. At this point, assuming the target contains sufficient detail, a match should have been found that allows the object to be recognized.

So, the generation of a 3D model description solves several problems inherent to object recognition. As the model is 3D, it allows recognition of the object from many angles and its hierarchical nature allows recognition of the entire object whilst maintaining more detailed information about the components.

3.2 Evaluating Marr and Nishihara's theory

Although it can be difficult to study the cognition involved in object recognition, there is evidence for some of the suggestions made by Marr and Nishihara.

One of the key predictions of their theory arises from the fact that they see establishing a central axis as a vital stage in the recognition process. This means that it should be very difficult to recognize an object if it is also difficult to establish the location of its central axis. Some support for this notion comes from a study conducted by Lawson and Humphreys (1996), in which participants had to recognize objects (line drawings in this case) that had been rotated. Rotation did not appear to have an effect on recognition unless the major axis of the object was tilted towards the observer. Presumably, the disruption to recognition was due to the major axis appearing foreshortened and therefore harder to locate.

More powerful evidence in support of Marr and Nishihara's theory comes from neuropsychological case studies. Warrington and Taylor (1978) reported that patients with damage to a particular part of the right hemisphere could recognize objects when they were presented in a typical view but not when presented in an unusual view. These patients also found it very difficult to say whether two photographs (presented simultaneously) were of the same object when one image was a typical view of that object and one an unusual view.

One explanation for this effect is that the patients could not transform the two-dimensional representation of the unusual view of the object into a 3D model description. However, as well as it being difficult to establish the central axis of an object presented in an unusual view, it is also likely that rotation would cause some key features of the object to become hidden. Humphreys and Riddoch (1984) prepared images of objects in which *either* a critical feature was obscured *or* the central axis had been foreshortened through rotation. These images were presented to patients similar to those tested by Warrington and Taylor. The patients had far more problems recognizing the axis-foreshortened objects than those with a key feature hidden. The results of these studies do offer some evidence that axis location may play a central role in generating a 3D model description of an object.

3.3 Biederman's theory

Marr and Nishihara's work has been extended and adapted in several related theories of object recognition. The most influential of these was proposed by Biederman in 1987. Biederman's theory (1987a) was also based on representing complex objects using a series of more simple primitives. Unlike Marr and Nishihara, Biederman did not restrict these primitives to generalized cones. Instead he proposed that the basic building blocks for describing an object were a set of basic shapes such as cylinders and cubes, known as **geons** (an abbreviated form of the phrase 'geometric ions'). Many of these geons are generalized cones, but they also include other

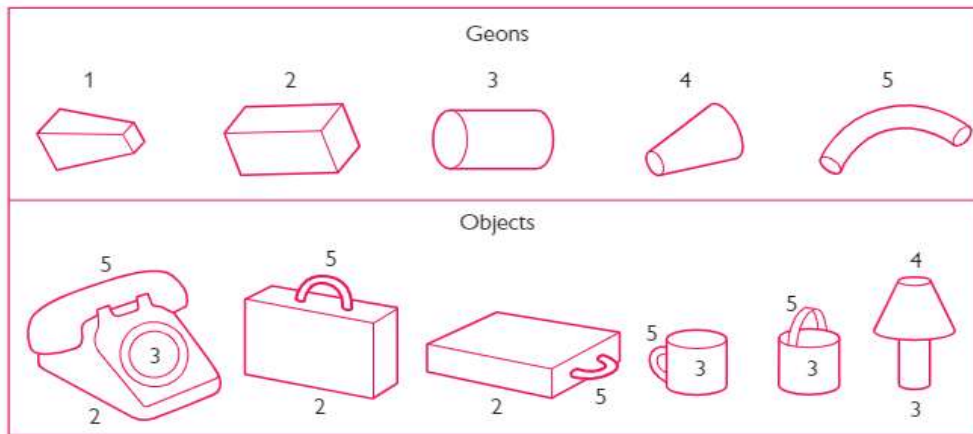


FIGURE 4.19 A selection of geons. Source: Biederman, 1987b

3D shapes that are very useful in representing common objects. A subset of geons is shown in the top part of Figure 4.19.

Biederman suggested that approximately 36 geons are needed in order to produce descriptions of all common objects. As with Marr and Nishihara's theory, more complex objects are represented by several different components and the division into components is based on areas of concavity.

The principal way in which Biederman's theory diverges from Marr and Nishihara's approach is the way in which a 3D description is formed from information in a 2D image – in other words, how the information in the primal sketch can be used to generate a 3D object-centred description. Biederman proposed that Marr's contour generators are not necessary to recover 3D shape, as each geon will have a key feature that remains invariant across different viewpoints. Thus, all that needs to be done is to locate these key features in the 2D primal sketch. Each feature can then be matched to a geon so that a 3D structural description of the object is generated. This description is then matched against those stored in memory.

Behind the concept of key features that remain invariant across viewpoint is the idea that some regular aspects of a 3D shape will tend to remain constant in any 2D image that is formed of that object. Biederman termed these 'non-accidental' properties to distinguish them from any regularity that was due simply to a particular viewpoint.

Biederman listed five non-accidental properties:

- *Curvilinearity* – a curve in the 2D image is produced by a curve on the object.
- *Parallelism* – lines that are parallel in the 2D image will be parallel on the object.
- *Cotermination* – two or more edges that terminate at the same point in the 2D image will terminate at the same point on the object.
- *Symmetry* – if the 2D image is symmetrical then the object will contain the same axis of symmetry.
- *Collinearity* – a straight line in the 2D image is caused by a straight line on the object.

Choosing which geon to use in order to represent an object (or part of an object) is then simply a matter of detecting these non-accidental properties and selecting a geon that shares them. For example, the 2D image of a ball will be a circle and will therefore contain no parallelism, cotermination, or collinearity, but will contain curvilinearity and an almost infinite degree of symmetry. The only geon to share these properties is a sphere, so the 3D shape of the ball is correctly described by a spherical geon.

Although these assumptions allow apparently ambiguous 2D images to be turned into an accurate 3D description, they can also lead to misinterpretation. For example, if you look at the wheel of a bicycle that is directly in front of you so that the wheel is viewed edge-on, its edges will appear to

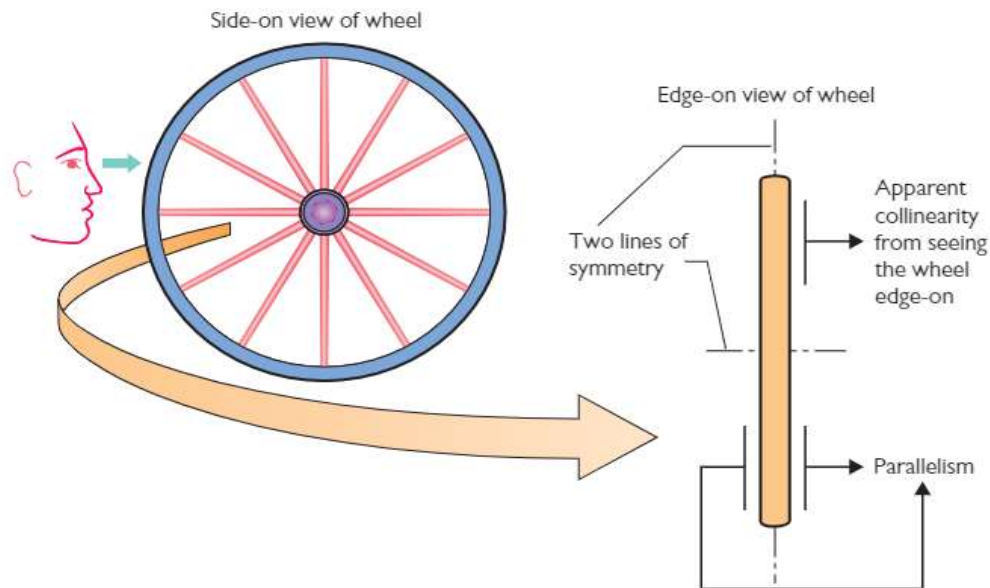


FIGURE 4.20 Apparent non-accidental properties of a wheel viewed edge-on.

have the following non-accidental properties (see Figure 4.20):

- *Collinearity* – the two vertical edges will appear as straight lines.
- *Symmetry* – there will be two lines of symmetry, one horizontal and one vertical.
- *Parallelism* – the two vertical edges will appear parallel.

However, the first of these non-accidental properties (collinearity) will be incorrect because a wheel does not contain any straight edges. We only see straight edges because of the viewpoint.

Although describing an object using non-accidental properties to select geons can lead to problems, there is evidence that supports Biederman's theory. The premise that concavities are used to divide the object into components (this premise was also used by Marr and Nishihara) was studied by presenting participants with images of objects that had part of their contours deleted. Deleting the part of the contour that corresponded to a concavity (that therefore occurred between components) resulted in a greater disruption to recognition than deleting part of the contour from elsewhere on the object (Biederman, 1987a).

3.4 Evaluating Biederman's theory and the object-centred approach

Representing objects through geometric ions or generalized cones may seem a little artificial to you, and indeed Biederman and colleagues wondered this themselves, and more specifically whether participants' sensitivity to the non-accidental properties of an object stemmed from being immersed in a manufactured world in which simple, regular shapes are very common. To test this, Biederman *et al.* (2009) compared the responses of participants drawn from two populations: students from the University of Southern California (USC) and members of the more remote tribes of the Himba, semi-nomadic people from Namibia. Both groups were asked to match a simple, regular geon to one of two other geons, which differed either according to a non-accidental property or a simple metric property. The results showed no significant difference between the responses of the Himba and the USC student participants, suggesting that the Himba are just as sensitive to the non-accidental properties of objects as are people accustomed to a more manufactured environment.

The production of an object description that is independent of viewpoint is a crucial stage in the theories of both Marr and Nishihara, and Biederman. So is there evidence that recognition does involve the generation of an object-centred description, rather than relying purely on viewer-centred descriptions?

To investigate the extent to which recognition is object-centred, Biederman and Gerhardstein (1993) used a technique known as **repetition priming**, where the presentation of one stimulus will make recognition of a related stimulus faster and/or more accurate. The idea behind their experiment was that if an object-centred description were being formed, then presenting one particular viewpoint of an object should facilitate (or prime) recognition of the same object presented in a different view. Their results showed that one viewpoint of an object did prime recognition of a separate viewpoint, as long as the change in the angle of viewpoint was not more than 135 degrees. However, even if the viewpoints were less than 135 degrees apart, if one or more geon(s) was hidden between the first and second view, then the amount of priming was reduced. This result supports both the idea that an object-centred description is generated (otherwise different viewpoints should not prime each other), and that this makes use of geons.

However, other researchers have reported results that do not appear consistent with Biederman and Gerhardstein's findings. Bulthoff and Edelman (1992) found that participants were generally unable to recognize complex objects that were presented in a novel viewpoint, even if the view of the object was one that should have allowed the generation of an object-centred description. Biederman and Cooper (2009) reflected on the priming work conducted since their

original experiments and pointed out that although visual priming does appear to be mediated by a representation of an object's shape that is independent of its position, size, or orientation in depth, it is also the case that participants form an episodic representation of the *specific* image originally shown. That both a view-independent *and* a view-dependent representation seem to be formed by exposure to one image obviously poses a challenge to theories of object recognition that do not incorporate both forms of representation.

In the end, then, it is unlikely that recognition is completely reliant upon the generation of object-centred descriptions such as those suggested by Marr and Nishihara (1978) and by Biederman (1987a), and there may well be tasks and occasions that involve viewpoint-dependent recognition, as suggested by Tarr (1995).

One task that is hard to incorporate into either Marr and Nishihara's or Biederman's theory is that of within-category discrimination. By representing objects as models consisting of either generalized cones or geons, a wealth of information is inevitably lost. For example, it is very likely that two collie-shaped canines would be represented as identical 3D models, yet it is possible to tell a border collie from a rough collie and even to tell specific dogs apart.

It makes sense that there should be more than one way of arriving at such a complex cognitive achievement as object recognition. In the theories we have examined in this section, the process of recognition has been conceived of as almost wholly passive and based on a single retinal 'snapshot' or view. As we have stated previously, there are different types of recognition and different ways of achieving it, including taking a more active approach.

SUMMARY OF SECTION 3

- Objects can be recognized from many different angles, suggesting that the process of recognition may be based on the generation of a 3D object-centred description.
- Marr and Nishihara (1978) suggested a theory of object recognition based on generating 3D models. This was achieved by: deriving the shape of an object from the 2½D sketch; dividing it into 'primitives' using areas of sharp concavity; generating an axis for each of these components; and representing each component via a generalized cone.

- The 3D models were hierarchical in nature and so included both global and detailed information stored in a hierarchically organized catalogue.
- Biederman (1987) suggested a similar theory based on using the non-accidental properties of an object to generate a description in terms of a series of basic volumetric forms known as geons.
- Although there is evidence that supports the approach taken by Marr and Nishihara and by Biederman, there are some forms of recognition that are difficult to explain using their theories.

4 FACE RECOGNITION

The act of recognizing the face of someone familiar is part of everyday life and most of the time it happens very quickly and effortlessly. Faces are arguably our favourite visual stimuli and can provide us with a great deal of very useful information. First and foremost we are very good at determining the identity of a face that is familiar to us, but faces also provide information about gender, age, and cultural background, as well as the emotional state of a person. Although it was only in the mid-1970s that psychologists acknowledged that faces were a unique visual stimulus, since then an enormous body of research on face recognition has been conducted, meaning that in this chapter we are only able to provide you with a snapshot of some of this research.

4.1 Comparing object and face recognition

Face recognition is very problematic for the 3D model approaches we have looked at so far. If we return to Humphreys and Bruce's model of object recognition shown in Figure 4.2, we can see that theories such as those of Marr and of Biederman have concentrated on the 'perceptual classification' stage of the process. Although this stage may provide information that is useful for navigation and basic interaction with the objects we find in the environment, more complex interaction is often necessary. For example, when you

are confronted by a person, you want to know not only that there is a human face in front of you, but *whose* face it is. This requires a much finer level of distinction than simply recognizing a sphere as a sphere; you must be able to tell which *specific* face is in front of you. As we shall see in Sections 4 to 7 of this chapter, the need to recognize individual faces has led to theories and research concentrating on different issues from that conducted within the area of more general object recognition.

Faces can be categorized at several different levels. At one level, we decide that the stimulus is a face as opposed to some other object. At another level, we decide that the face is female or male or derive other semantic information such as ethnic origin. We may even make attractiveness judgements. Importantly, we also decide whether the face is familiar or unfamiliar. If the face is familiar, there is also the need to decide to whom the face belongs and it is at this level that faces are rather different from other objects. It is this within-category judgement, which is like recognizing a specific cat or a specific cup, that sets face recognition apart from object recognition more generally, and it is regarded as more visually demanding because the differences between faces can be fairly minimal.

Tanaka (2001) has found evidence to liken this level of face recognition to expert recognition – for example, the expertise that certain individuals acquire through training in bird-watching or x-ray analysis. But whereas only some specifically trained

people achieve object expertise, face expertise is a general expertise that we all share and acquire without specific training. Whether or not this face expertise is the result of an innate processing system or the expression of a learned skill is a matter of debate and an issue we will return to in Section 7 of this chapter.

4.2 Recognizing familiar and unfamiliar faces

So how good are we at recognizing faces and identifying people? You already saw in Activity 4.2 that it was possible to recognize a face that was familiar to you (Paul McCartney) despite quite large changes in appearance. In fact, when you think about it, you are able to recognize your family and friends from any angle, under different lighting conditions, and even when they age or change their hairstyle, and you are still likely to be able to do this in 30 years' time. There is evidence to suggest that we can remember the names and faces of school-friends over long periods of time; recognition tests revealed hardly any forgetting over a 35-year period (Bahrck *et al.*, 1975). This is not the case with all the faces we encounter though. Later work by Bahrck investigated the ability of college teachers to recognize former students taught over a 10-week period (Bahrck, 1984). The teachers had met these students three to five times a week. Although the level of correct face recognition for those taught recently was reasonably high at 69 per cent, this dropped as the number of intervening years increased so that after 8 years only 26 per cent of the former students were correctly recognized.

What about faces that are not so familiar, for example ones we've only seen once or twice before? How well do we recognize unfamiliar faces? Indeed, that might sound an odd question, as how could it be possible to recognize a face you are unfamiliar with? Of course, it is impossible to recognize a face you have never seen before, but there is a great deal of difference between recognizing a familiar face (such as

those of family and friends) and recognizing a face we have only seen a few times. Most research on unfamiliar face recognition starts with a learning phase, where participants are presented with faces unknown to them that they then 'learn', and concludes with a test phase in which the participants attempt to recognize those faces they had seen previously from amongst other faces that they have never seen before. A large number of face-learning experiments have been conducted (e.g. Yin, 1969) and these have found that, when given an immediate recognition test, participants performed extremely well. (For example, Yin observed that participants correctly recognized 93 per cent of the faces previously shown to them). However, if the picture of the unfamiliar face shown in the recognition test depicted a different viewpoint or expression, then recognition rates dropped (e.g. Bruce, 1982), suggesting that what is being tested is 'recognition of a specific picture of a face' rather than 'face recognition' as we encounter it in everyday life. So, whilst we can recognize familiar faces independently of viewpoint or expression, the processing of unfamiliar faces is influenced by both viewpoint and expression. Furthermore, the environmental context in which the face is seen also has a larger influence on the recognition of unfamiliar faces than on the recognition of familiar faces (see Memon and Bruce, 1985, for a review). Finally, there is also considerable evidence from applied research on eyewitness identification decisions and from publicized cases of misidentification, such as the one involving Jean Charles De Menezes, that suggests unfamiliar face recognition is far, far more error prone than familiar face recognition.

In Box 4.1, you will see that research has also shown that even when participants do not have to recognize an unfamiliar face from memory, they are still prone to making errors.

The difference between familiar and unfamiliar face recognition might be quantitative in nature and simply reflect the amount of exposure the participants have had to the faces. However, according to Megreya and Burton (2006), it is more likely that the difference is qualitative in nature. They carried out six experiments involving face matching tasks

→ BOX 4.1 RESEARCH STUDY Recognizing unfamiliar faces in matching tasks

Even matching unfamiliar faces that are presented simultaneously (a task that does not test our memory) appears to be surprisingly difficult. In a field experiment, Kemp *et al.* (1997) looked at how well cashiers could match shoppers to credit cards bearing their photographs. They found that cashiers would frequently accept credit cards depicting a photograph of someone who bore a resemblance to the shopper (the correct decision rate to reject the card was only 36 per cent). Even when the photograph was of someone who bore no particular resemblance to the shopper but was of the same sex and ethnic background, the correct decision rate to reject the card was only 66 per cent (see Activity 4.4).

Other studies have demonstrated that we are not very good at matching two similar high quality photographic images when the face is unfamiliar. Bruce *et al.* (1999) showed participants a high quality video still of an unfamiliar young male target that was then presented in a line-up of similar images of nine other young men. Even when told that the target was definitely present in the line-up, participants picked it out accurately in only 80 per cent of the trials. If not told that the target was present, or if the pose of the target was varied between initial presentation and test, then performance was even worse. In fact, the performance of these participants has been matched or even exceeded by that of an automatic face recognition system tested on the same images (Burton *et al.*, 2001).

* ACTIVITY 4.4

Look at the images of three faces shown in Figure 4.21. Which of the images to the left (a or b) do you think is of the same woman as that in the right-hand image (c)? These images are examples of images that were used on photo credit cards in the study conducted by Kemp *et al.* (1997).

COMMENT

The correct answer is that the left-hand image (a) is of the same woman shown in (c), but cashiers often refused to accept it due to the change in hairstyle. However, the image in the centre (b) was often incorrectly accepted as being of the woman to the right (c).



(a)



(b)



(c)

FIGURE 4.21 Three faces.

similar to those described in Box 4.1. They presented the faces in both upright and inverted (upside down) conditions, and found that whilst performance for unfamiliar faces was similar in both conditions, participants were better with familiar faces in upright conditions. This finding was replicated in a further study (Megreya and Burton, 2007). Their conclusion was that unfamiliar faces are not processed for identity like familiar faces, and whereas unfamiliar face matching may rely on image-based or pictorial processes (where the faces are treated

as ‘simple visual patterns’), recognition of familiar faces is likely to involve a more specialized and robust type of processing.

The next section looks more closely at what recognition of familiar faces may entail. We will start our discussion by considering some of the errors people make. These errors provide us with important information about the different systems and processes that may be involved in face recognition. Importantly, models of face recognition need to be able to account for such errors.

SUMMARY OF SECTION 4

- Face recognition is an example of a within-category judgement task.
- Our ability to identify familiar faces is extremely good and relatively unaffected by pose, viewpoint, or context.
- Recognition of unfamiliar faces is much poorer and is influenced by changes in pose, viewpoint, or context.

5 MODELLING FACE RECOGNITION

The theories of object recognition we have looked at previously centred on matching the description of an object that is in view with a stored representation. Although face recognition also involves similar matching processing, this is not usually considered the end point. In addition to matching the face we also need to access relevant semantic information and, preferably, the person’s name.

5.1 Young *et al.*’s (1985) diary study

In a diary study, Young *et al.* (1985) asked 22 participants to make a record of the mistakes they made in recognizing people over an eight-week period. The recorded errors or difficulties tended to fall into different categories as shown in Table 4.2.

* ACTIVITY 4.5

Although we may have face recognition expertise, we do make mistakes. Before reading on, reflect for a moment and recall the last time you discovered

that you failed to recognize someone you know or you mistakenly thought you recognized someone you didn’t know.

TABLE 4.2 The main types of everyday errors in face recognition revealed by Young *et al.* (1985)

Types of everyday errors	Number of errors
Person misidentified	314
Person unrecognized	114
Person seemed familiar only	223
Difficulty in retrieving full details of the person	190
Decision problems	35

What do these different categories mean? ‘*Person misidentified*’ refers to those occasions when someone unfamiliar is misidentified as someone familiar and ‘*Person unrecognized*’ refers to occasions when someone familiar was thought to be someone unfamiliar. Both may arise because of poor viewing conditions (i.e. it is a bit dark) or because we know the person only slightly. ‘*Person seemed familiar only*’ refers to those occasions when you recognize someone as being familiar but no other information comes to mind immediately, and ‘*Difficulty in retrieving full details of the person*’ refers to occasions when only some semantic information is retrieved and not, for example, their name. These errors often occur when the familiar person is seen outside the context in which they are usually encountered. Finally, ‘*Decision problems*’ refers to those occasions where you think you recognize the person but decide it cannot be them, perhaps because you believe they are currently in another country.

These errors demonstrate that it is possible to retrieve previously learned semantic information about a person without recalling their name, but that it is very unlikely that we will recall their name without also retrieving relevant semantic information. However, before we can recall either semantic information or a name, we must realize the face is familiar.

5.2 A cognitive model of face recognition

These findings on everyday errors are consistent with the notion that the recognition of familiar faces

involves a sequence of processes using different types of information. Hay and Young (1982), Young *et al.* (1985), and then Bruce and Young (1986) refined a cognitive theoretical framework or model of person recognition involving such a sequence of stages. On meeting people we encode their faces. This encoded information may activate **face recognition units** (FRUs) that contain stored information about the faces we are familiar with. If there is a reasonable match between what has been encoded and what is stored in the recognition unit, then the recognition unit will be activated and allow access to semantic information about the person’s identity, such as their occupation, stored in **person identity nodes** (PINs). It is only once the PIN for a face has been activated that their name can be generated. A **cognitive system** is also involved, as the information provided by the recognition system must be evaluated. As the diary study above indicated, errors in face recognition can arise because of decision problems. For example, if we know that the person doesn’t live or work nearby, that knowledge may override what our recognition system is telling us and hence we may doubt that we have correctly identified the person.

So, when might the encoded information activate FRUs? As mentioned in Section 4, familiar face recognition is extremely good and relatively unaffected by pose, viewpoint, or context, and to achieve this Bruce and Young suggested that structural codes are formed over time. These codes contain the information that allows a face to be recognized despite changes in pose, expression, etc. Importantly, structural codes allow us to decide whether or not a face is familiar. The first stage of the model involves the creation of these codes resulting in two types of representations that serve different purposes. Viewer-centred descriptions result in viewpoint-specific representations and, as can be seen in Figure 4.22, these are used for expression analysis and to help make sense of speech. Expression-independent descriptions result in separate and more abstract expression-independent representations, which are used for recognition. Therefore, a face needs to be familiar if it is to activate a FRU; if it is unfamiliar then its representation will be linked to the particular view of the face that is being shown in that particular encounter.

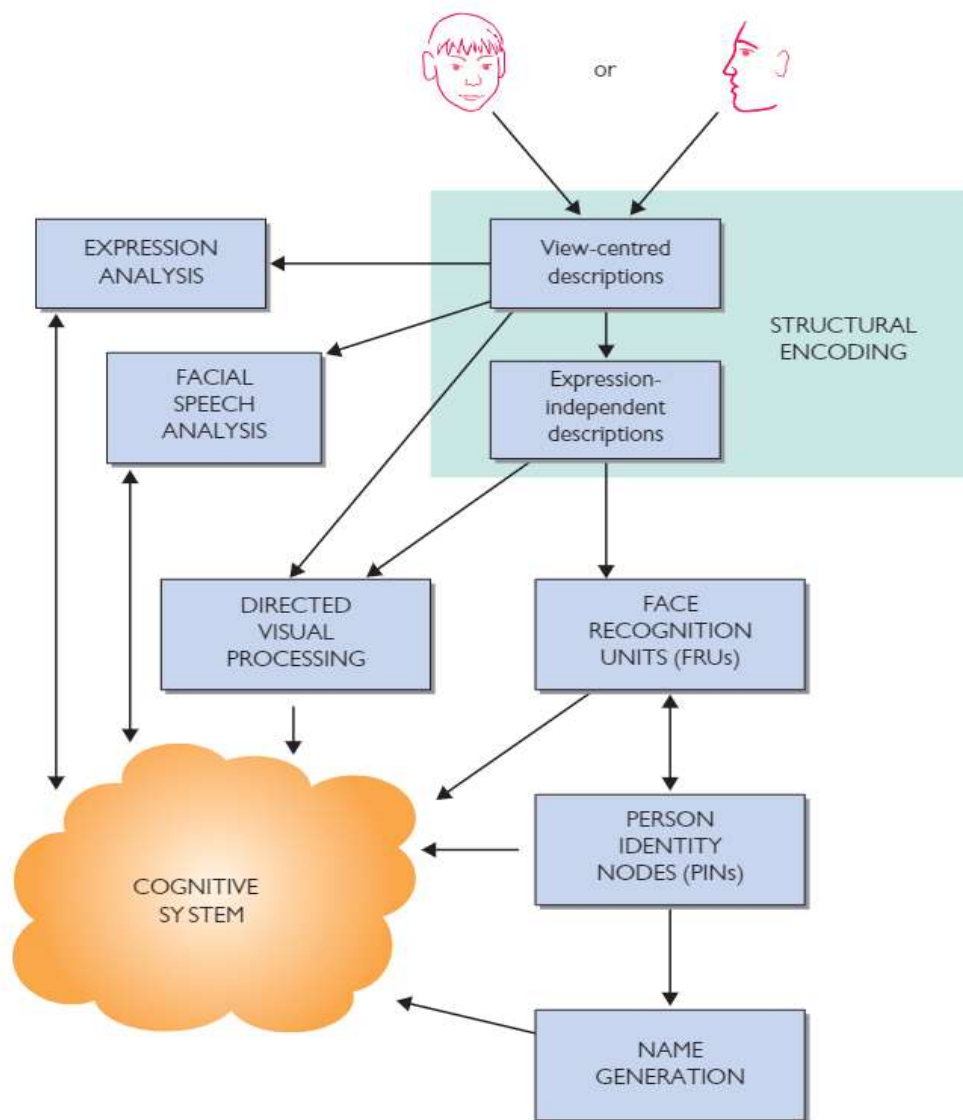


FIGURE 4.22 Bruce and Young's functional model for face recognition. Source: Bruce and Young, 1986, p.312

The Bruce and Young (1986) functional model for face recognition is presented in Figure 4.22. As you can see, there are separate routes for facial expression analysis, facial speech analysis, and face recognition; and face recognition progresses through a sequence of stages from FRUs to PINs to name generation.

The notion that different types of information are sequentially accessed is also supported by the results of experiments conducted in the laboratory. For example, Hay *et al.* (1991) showed participants 190 images of the faces of celebrities (including people likely to be familiar to the participants and those likely to be unfamiliar) and asked them to decide whether or not each face was familiar and to state the person's occupation and name. Participants did not retrieve a

name without also being able to name the occupation, thus supporting the notion that semantic 'person identity' information is retrieved *before* the person's name. Other studies (e.g. Johnston and Bruce, 1990), looking at how quickly we can complete a particular task, have shown that faces can be classified as familiar more quickly than they can be classified by occupation, and furthermore that classifications that require accessing the person's name take longer than classifications involving a person's occupation or other semantic properties. These findings support the notion that perceptual classification, judging the familiarity of a person, takes place *before* semantic classification and that a person's name is accessed last. They also provide a nice demonstration of how the

findings from the laboratory may support those derived in a more everyday study of face recognition, such as Young *et al.*'s (1985) diary study.

5.3 A connectionist model of face recognition

The IAC model (e.g. Burton *et al.*, 1990; Burton and Bruce, 1993) is a connectionist model (recall the discussion of connectionism in Chapter 1) of face

recognition and an extension and implementation of the Bruce and Young model described above. IAC stands for 'interactive activation and competition network'. As this model is a computer simulation of face recognition it has been tested by seeing how compatible it is with the available evidence, and by looking at the predictions it generates.

The model comprises units that are organized into pools (see Figure 4.23). These pools contain:

- *FRUs (face recognition units)*: For every familiar person, there is one FRU in the model. These are

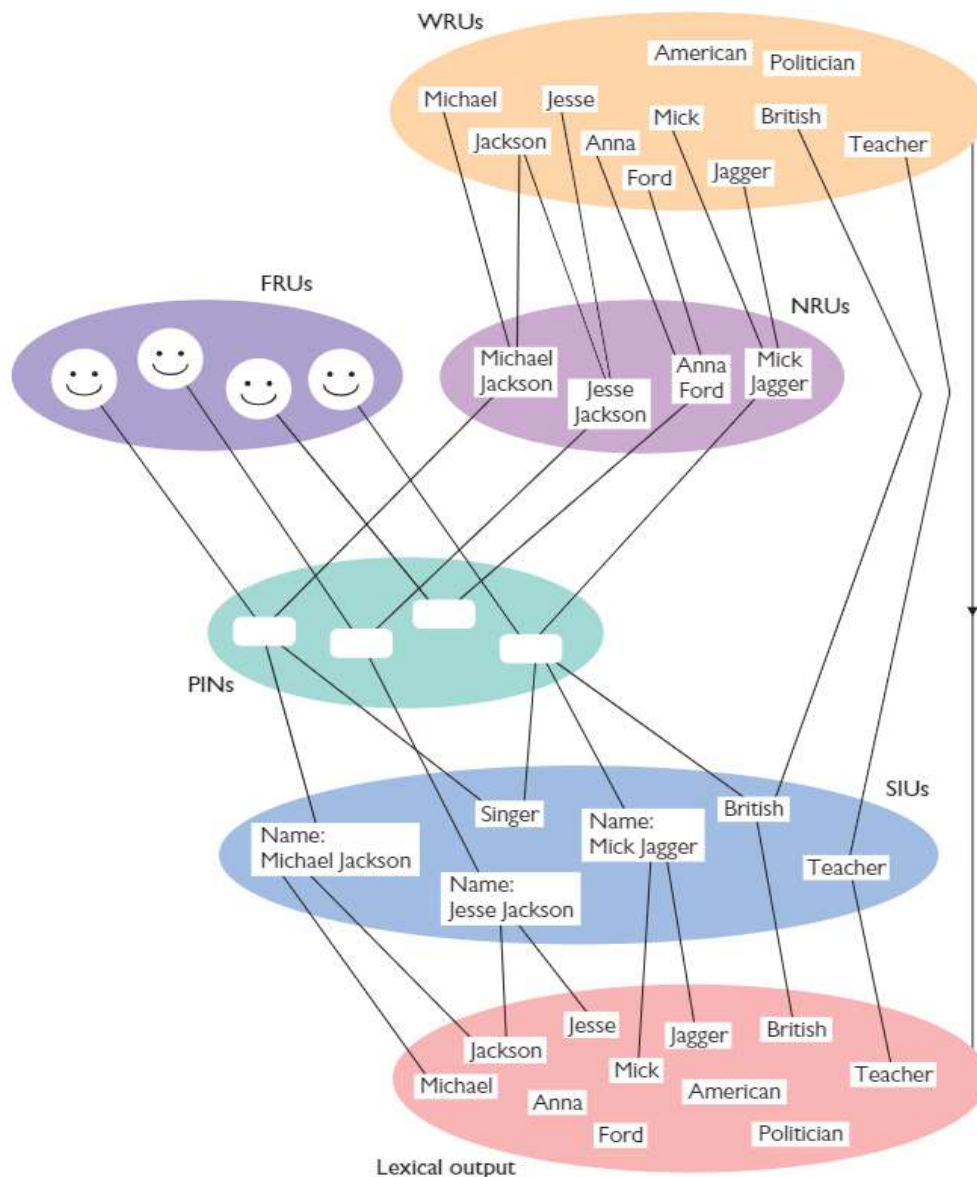


FIGURE 4.23 The central architecture of the IAC model.

view-independent and seeing any recognizable view of a face will activate the appropriate FRU. These representations allow perceptual information to be mapped onto stored memories. (This is basically what was suggested in the Bruce and Young model.)

- *PINs (person identity nodes)*: This is where a face is classified as belonging to a person, and there is one unit per known person.
- *SIUs (semantic information units)*: Relevant semantic information is stored here, e.g. occupational category.
- *Lexical output*: Units representing output as either words or a name.

The IAC model also includes a route based on word recognition. The pool of WRUs (word recognition units) represents an input lexicon containing both specific names and more general information, such as nationality or occupation. Words that are names have direct links to a pool of NRUs (name recognition units), which are linked to PINs in the same way as FRUs. The WRUs that do not correspond to names are linked to SIUs.

Figure 4.23 shows how the pools are connected. The input systems (FRUs and NRUs) join to a common set of person identity nodes (PINs) and these are linked to units containing semantic information (SIUs). Each of the pools is illustrated here with just a few examples of the units they might contain. Many SIUs will be shared and here many people will be represented with such information as ‘teacher’ or ‘British’.

Recognizing a face is modelled in the following way: seeing a face will activate an FRU, which in turn increases activation in the relevant PIN. As PINs are linked to SIUs, activation of the PIN will bring about activation in the relevant SIU. The notion that different types of information are sequentially accessed is therefore still present in this connectionist model. If a certain threshold is achieved in the PIN, then this signals familiarity. An important point to note is that different types of information come together at the PIN stage, including information from recognition systems

specialized for faces, as well as those specialized for the recognition of written or spoken names, and familiarity is judged on this pooled information.

We mentioned before that IAC stood for ‘interactive activation and competition network’. The ‘interactive activation’ arises from the links between units in different pools, which are *excitatory*: the FRU for Mick Jagger’s face excites or activates the PIN for Mick Jagger, which in turn excites semantic information units for the name ‘Mick Jagger’, the occupation ‘singer’, and the nationality ‘British’. These excitatory links are bidirectional so that excitation also runs in the opposite direction from ‘singer’ to Mick Jagger’s PIN and Mick Jagger’s FRU. However, within each pool links between units are *inhibitory* (these links are not shown in Figure 4.23), so this is where ‘competition’ arises. Excitement in the FRU for Mick Jagger will inhibit activity in the other FRUs, just as excitement in Mick Jagger’s PIN will inhibit activity in the other PINs and excitement in one SIU will inhibit activity in another SIU. But the SIU for Mick Jagger (which might be ‘singer’) will also excite many other PINs (in this example, those belonging to all other singers). This means that activation of PINs will not be limited solely to the specific person in question, but some activation will also occur for anyone who is semantically related (e.g. shares the same occupation). Thus, the model incorporates the results of experiments that have shown priming effects – that you are quicker to recognize Bill Wyman if you have already seen Mick Jagger. Generally, the strength of this connectionist model is that it can account for findings from laboratory studies as well as for the everyday errors described by Young *et al.* (1985).

To conclude this section, it is important to highlight that the Bruce and Young model has been very influential and still remains so today, providing a framework for organizing the different elements involved in face recognition. It has been superseded by the IAC model, which provides more information on how FRUs and PINs actually operate and accounts for the effects of priming. Next we shall consider how well the IAC model can account for findings from neuropsychological studies.

SUMMARY OF SECTION 5

- Everyday errors suggest that recognizing faces involves sequential access to different types of information.
- A cognitive model of person recognition involving such an idealized sequence of stages has been developed (Bruce and Young, 1986).
- IAC is a connectionist model of face recognition that is an extension and implementation of this cognitive model.

6 NEUROPSYCHOLOGICAL EVIDENCE

Prosopagnosia, the inability to recognize faces whilst maintaining the ability to recognize other objects, is a well-documented phenomenon. However, cases of 'pure' prosopagnosia are exceptionally rare. It is more common to see deficits affecting other visual categories too. The recognition of all familiar faces is affected, regardless of their semantic categories (so it is not the case that the failure to recognize a face is restricted to faces of celebrities or politicians). However, as recognition from other cues, for example voice, usually remains unaffected, the condition is specific to visual recognition of faces and is not a more general impairment of the recognition of personal identity. Also, the ability to distinguish between faces is often preserved.

In this section, we shall focus on two key findings that have emerged from investigations of prosopagnosia: first, that identification of expression appears to be independent from face identification; and second, that face recognition and awareness of face recognition might also be independent of one another. It is possible that although prosopagnosics are unable to recognize faces consciously or overtly, certain types of non-conscious response may be preserved. We shall examine how the IAC model may account for this.

6.1 Separate routes for identification and emotional expression

As mentioned in Section 5, models of face recognition have proposed a route for face identification that is independent of emotional expression, and this independence has received support from experimental work and from neuropsychological research. In many cases of prosopagnosia, the ability to recognize facial expressions may be unaffected. Young *et al.* (1993) looked at ex-servicemen with unilateral brain injuries and tested familiar face recognition, unfamiliar face matching, and analysis of emotional facial expressions. Analysis of accuracy data showed evidence of selective impairments in each of these three abilities. For example, one participant with a right hemisphere lesion was selectively impaired in identifying familiar faces, whereas a different participant, also with a right hemisphere lesion, had problems only with matching unfamiliar faces. A number of other participants with left hemisphere lesions were only impaired on the facial expression tasks. Response latency data also supported the notion of a selective deficit of facial expression processing, but suggested that

impairments of familiar face recognition and unfamiliar face matching were not entirely independent from one another. The findings from this study thus provide strong support for the notion that facial expression analysis and face identification seem to proceed independently of each other (and also some support for the notion that the ability to recognize familiar faces and to match unfamiliar faces may be selectively and independently impaired).

6.2 Covert face recognition

Previously, when describing models of face recognition, we did not draw a distinction between face recognition and awareness of recognition. However, neuropsychological research on prosopagnosia suggests that the distinction is important. Bauer (1984) monitored changes in autonomic nervous system activity via changes in **skin conductance response (SCR)**. These changes signal an affective or emotional reaction (you may remember reading in Chapter 2 on attention how a closely related response, GSR, was measured to look at unconscious processes). Bauer showed LF, a participant with prosopagnosia, a face and read out a list of five names, whilst simultaneously measuring SCR. If LF was asked to pick the correct name he performed no better than expected by chance. In other words, LF was overtly unable to recognize familiar people from their faces. However, LF showed a greater SCR when the correct name was read aloud compared with the incorrect names. Thus, LF was showing an affective or emotional response, but this response was not a conscious one. The term **covert recognition** is used to describe this non-conscious recognition or emotional response to the faces.

Since Bauer's work, many studies have investigated covert recognition and the issue is not whether this type of face recognition exists but how to interpret it. Bauer proposed that separate neural pathways are responsible for two independent routes to recognition, one for conscious overt recognition and one for non-conscious covert recognition. Although questions remain over exactly how overt and covert recognition

processes are mediated and how these processes normally become integrated, there is support for the involvement of the two major neural pathways (see Box 4.2).

A different issue is whether those individuals who retain covert recognition can be helped to overcome their disorder. Could covert recognition be turned into overt recognition? Sergent and Poncet (1990) were the first to demonstrate such provoked overt recognition. In their study, PV was shown eight faces of famous people from the same semantic category and she was unable to identify them. However, when she was told that they all had the same occupation and she looked at the faces again, she was able to say that they were all politicians, name seven of the people, and recall biographical information about the eighth person. This and other later studies (e.g. Diamond *et al.*, 1994) have shown that provoked overt recognition can occur under certain experimental conditions, and this provides some hope for rehabilitative work.

Can the IAC model accommodate the pattern of deficits described here? Covert without overt recognition is explained in terms of attenuation (or weakening) in the connections between the FRUs and PINs. This means that when a face is seen and the FRU is activated, the weakened FRU-PIN connection strength means that excitation of the corresponding PIN is not raised above the threshold for the face to be recognized overtly. However, this weakened activation may be sufficient to raise the excitation of the PIN above its resting level, mediating covert recognition. Provoked overt recognition is explained in the following way. Telling PV that the faces are related is equivalent to strengthening the PIN-SIU connections. Unlike FRU-PIN connections, PIN-SIU links are assumed to remain intact in instances of prosopagnosia where covert recognition is observed. Once these connections are strengthened, activation is passed back from the shared SIUs to the relevant PINs. These then achieve threshold and the faces are recognized overtly. Simulations with the model confirmed this particular prediction – provoked overt recognition was successfully modelled (Morrison *et al.*, 2001).

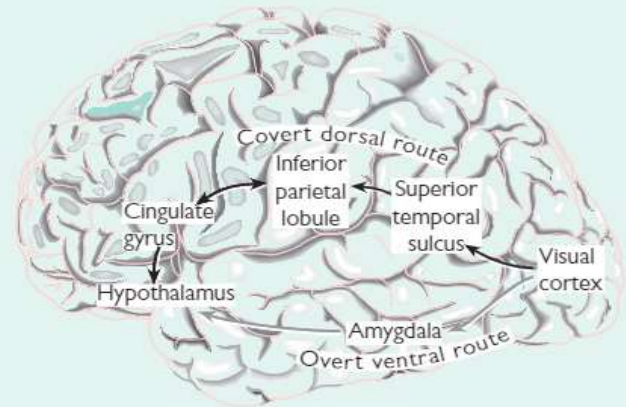
→ BOX 4.2 RESEARCH STUDY Capgras delusion

Capgras delusion usually occurs as part of a psychiatric illness, although it can result from brain injury. A person with Capgras delusion believes firmly that someone they know, usually a relative or close friend, has been replaced by an impostor, double, robot, or alien. Sometimes the delusion relates to objects; for example, the sufferer may believe that tools, ornaments, or other household objects have been replaced by doubles. Face and object Capgras delusion do not usually co-exist, and the disorder tends to be specific to one domain. The key point here is that individuals with a face Capgras delusion recognize a face but simultaneously refute its authenticity. Exactly why those with Capgras delusion adhere to the belief that the person must be an impostor is still being debated.

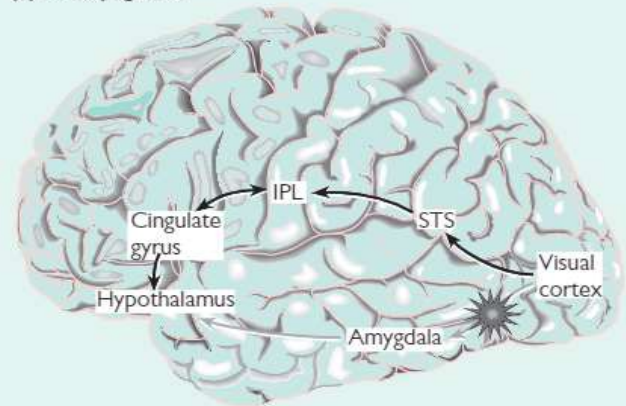
One of several explanations was offered by Ellis and Young (1990). They suggested that Capgras delusion may be a 'mirror image' of the impairments underlying prosopagnosia. Bauer (1984) proposed that the neuroanatomical pathway involved in overt recognition was the 'ventral visual-limbic pathway', whereas the pathway involved in covert recognition was the 'dorsal visual-limbic pathway'. Ellis and Young suggested that the Capgras delusion resulted from damage to such a dorsal route, so that sufferers would recognize the familiar person but not receive supporting affective information. Their prediction that individuals with Capgras delusion would recognize familiar faces but would fail to show an autonomic emotional response to these familiar faces has received support from several studies (e.g. Hirstein and Ramachandran, 1997). Whilst overt recognition is intact, covert recognition seems to be impaired.

Figure 4.24 shows normal face processing (a), with a darker arrow showing the covert dorsal route and a lighter arrow the overt ventral route. In prosopagnosia (b) the overt ventral route is thought to be damaged, and in Capgras delusion (c) the covert dorsal route is thought to be damaged.

(a) Normal face processing



(b) Prosopagnosia



(c) Capgras delusion

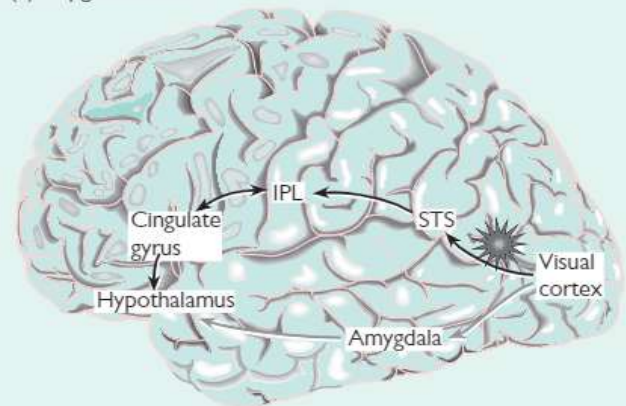


FIGURE 4.24 The dorsal and ventral routes in normal face processing (a), prosopagnosia (b), and Capgras delusion (c). Source: Ellis and Lewis, 2001, Figure 3, p.154

6.3 Evaluating the IAC model

As a model of face recognition, the IAC model is impressive in that it can account for a wide range of data from studies on face recognition. Whilst there are other models of face recognition, some of these are based on a narrower range of evidence; for example, they may have sought only to account for the findings from neuropsychological studies. As we have seen here, IAC is compatible with everyday, laboratory, and neuropsychological findings.

Hole and Bourne (2010) suggest that the weakness of IAC is that it says little about what information is extracted from the face in the first place to allow it to become familiar. How is a face encoded to allow it to be matched to an internal representation so that FRU activation then follows? One issue that has been raised is whether structural codes are indeed used to

combine the descriptions formed by multiple encounters with a (familiar) face, so that what is stored in memory is an abstracted representation of the face. Longmore *et al.* (2008) have challenged such a notion after carrying out six experiments where participants learned to recognize new faces to near 100% accuracy (they termed this familiarized face recognition). Their findings revealed that when participants learned multiple views of the same face, their performance decreased when shown novel views of that face, suggesting that the different views of the same face were stored as a collection of separate images rather than as a combined representation. Therefore, what was stored appeared to be based on view-specific rather than structural codes. We shall consider other findings regarding the information that is extracted from a face in the final section of this chapter.

SUMMARY OF SECTION 6

- Prosopagnosia is the inability to recognize faces, although expressions and other objects may still be correctly identified.
- Covert face recognition, shown by autonomic responses to faces, may however be spared.
- Overt conscious face recognition and covert non-conscious face recognition are different types of face recognition that may be mediated by different neural pathways.
- Capgras delusion may be a mirror image of prosopagnosia in terms of which system remains intact and which system is damaged.
- Provoked overt recognition has been achieved in some studies and has been successfully modelled using the IAC model.

7 ARE FACES ‘SPECIAL’?

In this last section we return to the issue of the difference between face recognition and object recognition, and in particular to face expertise and how we are able to discriminate so readily between faces. There are several important issues that the literature has addressed:

1. Is there a neuroanatomical location that underlies face processing and, if so, does this mean that face

processing is unique and qualitatively different from the processing of other types of visual stimuli?

2. Is face processing an innate or learned skill? Have we developed a face expertise because of constant exposure to faces and practice at differentiating between them, or is this an innate ability?

3. How important are the individual features of the face, the relationships between the features, or the three-dimensional structure? Do we process the individual facial features or the face as a whole?

In the previous section, we looked at the syndrome of prosopagnosia and found that research implicated several neurological pathways. Of particular interest is that prosopagnosia can leave object recognition relatively intact and, in turn, face recognition has been spared in cases where object recognition has been impaired (a double dissociation). Studies using the technique of functional magnetic resonance imaging (fMRI) have found facial stimuli to activate an area in the fusiform gyrus in the posterior temporal lobes (especially in the right hemisphere), whilst non-face objects activated a different area. There is also the observation of cells specialized for faces within the monkey temporal lobe – these cells respond selectively to faces of humans and/or monkeys but not to other stimuli (e.g. geometrical shapes and bananas). There is, therefore, evidence to suggest that the processing of faces is mediated by specific areas of the brain; that there is cortical specialization for faces. But does this mean that face recognition is *unique*, that the processes used for recognizing faces are qualitatively different from those used for recognizing other visual stimuli?

There is support for the notion that there is a special mechanism from birth for processing facial information, as newborn babies show a preference for face-like visual patterns. Rather than an innate neural mechanism that processes faces, Johnson and Morton (1991) suggested that there is a mechanism that makes newborns attentive to faces, and this innate attentional bias then ensures that any system for learning visual stimuli receives a lot of face input and learns about the individual characteristics of faces. Although there is a ‘kick-start’ mechanism that gives face processing in newborns a special status, which serves to guide subsequent learning, soon other processing systems will come into play that may or may not be unique to faces.

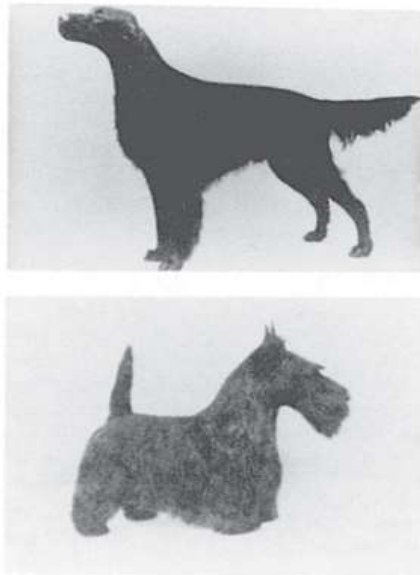
One further reason to think that face recognition is a special type of recognition, distinct from other object recognition, is that faces all tend to look alike in that they have similar features in similar positions. Given this similarity, it could be that we have to make

use of a different form of visual information to recognize a face from that used to recognize, for example, a table. Some evidence that this is indeed the case comes from studies that have demonstrated that inverting, or turning upside down, visual stimuli disproportionately impairs our ability to recognize faces compared with our ability to recognize objects. This is known as the *inversion effect*. Yin (1969) and other studies since (e.g. Johnston *et al.*, 1992) have shown that inverting a photograph of a face disrupts recognition more than does inverting an image of an object. Yin looked at the influence of inversion on faces and other stimulus material including houses and aeroplanes. Although recognition memory was better for upright faces than for other material, when the stimuli were turned upside down, recognition for faces was worse than that for other material. The key question is whether this peculiar reversal of recognition accuracy for faces (from best upright to worst inverted) supports the notion that faces are processed differently from other stimuli or whether there is an alternative explanation.

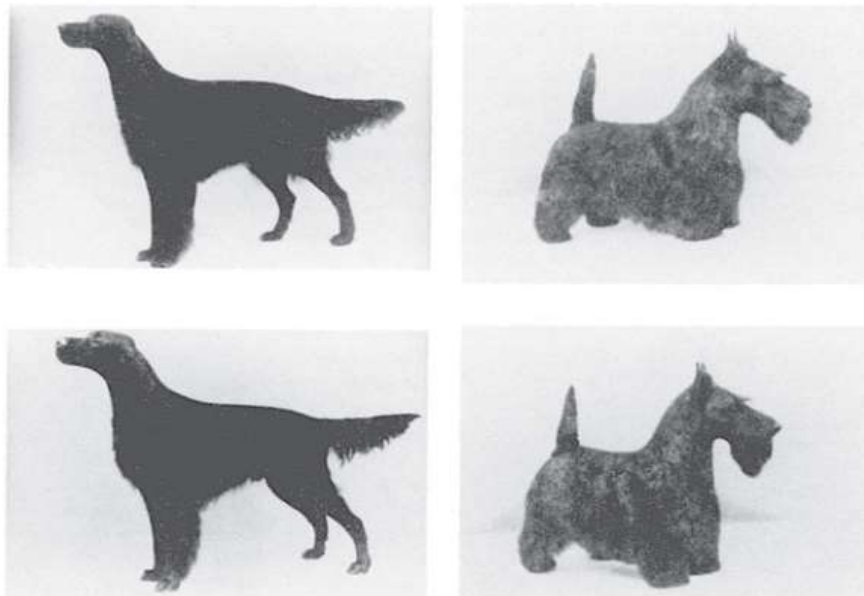
Diamond and Carey (1986) investigated an alternative hypothesis, namely that the effect of inversion on faces was a result of our perceptual mechanisms becoming ‘tuned’ to seeing this special type of visual stimulus in the usual upright orientation. This ‘tuning’ or expertise would then be ‘lost’ when we see them inverted. Their research considered whether the inversion effect was indeed specific to human faces or whether it would in fact arise when using any class of visual stimulus with which we have a large amount of experience. To investigate this, Diamond and Carey selected participants to include both people who were not interested in dogs and people who were dog experts (mainly dog-show judges, breeders/handlers, or people with a sustained interest in dogs). These participants were shown photographs of both human faces and dogs (body profiles – see Figure 4.25) and told to look at each photograph and try to remember it. Analysis revealed that whereas all participants recognized upright faces better than inverted faces, dog experts also recognized upright dogs better than inverted dogs. This finding has been interpreted as supporting the notion that the inversion effect is acquired as a result of expertise and is not a ‘face-specific’ effect.

What changes then in the way we process faces as we acquire this expertise? Diamond and Carey proposed a distinction between first-order and second-order relational properties. *First-order relational properties* refer to the spatial relationships among parts of the face; for example, the eyes are above the nose and the mouth is below the nose. Faces cannot be distinguished according to their first-order relational properties as they all

share the same basic arrangement or configuration. However, first-order relational properties help us detect that a visual stimulus is a face – a necessary step before identifying the face. *Second-order relational properties* refer to the differences in this basic configuration. This refers to the differences in the way the features are arranged in relation to each other; for example, wide-set eyes with a low forehead versus



(a) Inspection items that participants were asked to remember



(b) Recognition items: participants were asked to judge which of the stimulus items they had seen previously

FIGURE 4.25 Examples of the dog stimuli used. Source: Diamond and Carey, 1986, p.112

narrow-set eyes and a high forehead. Expertise results in a greater sensitivity to these second-order relational properties, as it is these properties that individuate members of the same class, such as human faces.

There is support for the notion that inversion influences our sensitivity to second-order relational properties. For example, Searcy and Bartlett (1996) presented participants with photographs of grotesque-looking faces. They created images where they had either distorted individual facial features (eyes and mouths) or they had distorted the spatial relations between the features (see Figure 4.26). They then presented these manipulated images in upright and inverted orientations. Participants rated the grotesqueness of the images and results showed that images of faces with distortions to the spatial relations between the features were rated as less grotesque when presented

inverted rather than upright; inversion failed to reduce ratings of grotesqueness when the distortions were performed on the features. These findings support the notion that inversion disrupts our processing of spatial relationships between the features.

Research like this suggests that our expertise in (upright) face recognition stems from the way in which these upright faces are processed as 'configurations' rather than as an assemblage of independent features. The term **configural processing** has been used, although this has been interpreted in a number of ways: to refer to the spatial relationships between features (i.e. second-order relational properties); to refer to the way facial features interact with one another (i.e. the way the shape of the mouth influences how the shape of the nose is perceived); to refer to holistic processing of the face (i.e. the face is perceived



FIGURE 4.26 Examples of stimuli used by Searcy and Bartlett (1996): the pair labelled 'A' shows a normal image and one with distorted facial features; the pair labelled 'B' shows a normal image and one with spatial distortion. Source: Searcy and Bartlett, 1986, Figure 1, p.907

as a whole face pattern and not broken down into separate features); or even to refer to the basic arrangement of the facial features (i.e. first-order relational properties). A considerable amount of research has been devoted to investigating the relative importance of this type of processing as compared with the processing of the facial features (known as featural processing or piecemeal processing). Although it is not always clear what different researchers mean by the term 'configural', there is agreement that configural information plays an important role in the perception and representation of upright faces.

One area of debate centres on the suggestion that this reliance on configural processing is the result of learning to recognize lots of faces, and hence the result of expertise. Although this does not rule out input from an innate mechanism, which may have 'kick-started' this learning by biasing attention towards faces, it does not suggest that face perception and recognition involve *unique* processes that are qualitatively different from those used to process other types of stimuli. This generic expertise hypothesis has been criticized by Robbins and McKone (2007), who failed to replicate the finding of Diamond and Carey (1986).

Robbins and McKone tested dog experts who they confirmed had good individuation of exemplars of their breed-of-expertise. Their experts included judges, breeders, and trainers of Labradors and so their stimuli were of Labrador Retriever dogs. They compared the performance of these experts with novices on a number of tasks, including inversion, and in contrast to Diamond and Carey found no difference between experts and novices for the dog-inversion effect. Their paper not only presented new data but also provided a review of behavioural experiments often used as support for the expertise hypothesis, and they concluded that configural/holistic processing does not occur for objects of expertise and is therefore unique to face processing. They claimed that the findings of investigations looking at configural processing support the idea of domain-specificity for faces; in other words that faces *are* special.

In sum, there is physiological and neuropsychological evidence supporting the existence of areas specialized for processing faces, and evidence suggesting an innate ability to pay attention to faces. Whether or not the processes involved in face recognition are unique is a central question under investigation.

SUMMARY OF SECTION 7

- Neuropsychological and physiological evidence suggests that there are specific areas of the brain that mediate face processing.
- Research on newborn babies suggests an innate ability to attend to faces.
- The inversion effect appears to be linked to our expertise in processing upright faces using configural information.
- We may develop expertise at distinguishing members of other categories of visual stimuli that also involves configural processing.
- Most evidence does not suggest that the processes involved in the perception and recognition of faces are unique.

8 CONCLUSION

In this chapter we have explored different types of recognition and looked at some of the mechanisms that

allow us to recognize objects and faces. In reading about recognition, you may well have got the idea that

cognitive psychologists still have a lot to learn about how object and face recognition may occur. Indeed, research has yet to clarify the different processes involved in recognizing familiar faces as opposed to unfamiliar faces, or to specify fully the overlap, if any, between the processes involved in face identification and those used in object recognition.

It is undoubtedly the case that a great deal of research is still being conducted in order to provide a

more comprehensive and detailed theory of the cognition involved in recognition. Just as there are different types of recognition, there are also different ways of recognizing faces and objects – for example, visually or by touch – and these different ways may involve different processes. So, rather than seeing the theories discussed here as providing a final answer, the best way to view them is as taking some of the initial steps in this complex but interesting field.

FURTHER READING

Bruce, V., Green, P.R., and Georgeson, M.A. (2003) *Visual Perception: Physiology, Psychology and Ecology*, Hove, Psychology Press. This textbook on visual perception by Bruce, Green, and Georgeson covers both perception and recognition and provides a more extensive look at the areas of cognition covered in both this and the preceding chapter.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco, W.H. Freeman & Company. The critical importance of the work of Marr for theories of perception and recognition is hopefully clear by now. Marr's seminal book, published in 1982, is still widely available and the best way of really getting to grips with his research.

Hole, G. and Bourne, V. (2010) *Face Processing: Psychological, Neuropsychological, and Applied Perspectives*, Oxford, Oxford University Press. This book by Hole and Bourne is an excellent text suitable for the student reader, that provides a comprehensive review of different areas of research on face processing. As well as covering empirical findings from laboratory based research, evidence from neuropsychological investigations is described and relevant applied research is discussed. It also covers theories of face perception and recognition that we have not had time to go into here, most notably Valentine's multi-dimensional face space model and the approach to face recognition based on principal component analysis.

REFERENCES

Bahrick, H.P. (1984) 'Memory for people', in Harris, J.E. and Morris, P.E. (eds.) *Everyday Memory, Actions and Absent-mindedness*, Academic Press, London.

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

Bauer, R.M. (1984) 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia*, vol.22, pp.457–69.

Biederman, I. (1987a) 'Recognition by components: a theory of human image understanding', *Psychological Review*, vol.94, pp.115–47.

Biederman, I. (1987b) 'Matching image edges to object memory', *Proceedings of the First International*

Conference on Computer Vision, IEEE Computer Society, pp.384–92.

Biederman, I. and Cooper, E.E. (2009) 'Translational and reflectional priming invariance: a retrospective', *Perception*, vol.38, pp.809–825.

Biederman, I. and Gerhardstein, P.C. (1993) 'Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance', *Journal of Experimental Psychology: Human Perception and Performance*, vol.19, pp.1162–82.

Biederman, I., Yue, X., and Davidoff, J. (2009) 'Representation of shape in individuals from a culture with minimal exposure to regular, simple artifacts: sensitivity to nonaccidental vs. metric properties', *Psychological Science*, vol.20, pp.1437–42.

- Bruce, V. (1982) 'Changing faces: visual and non-visual coding processes in face recognition', *British Journal of Psychology*, vol.73, pp.105–16.
- Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M., and Miller, P. (1999) 'Verification of face identities from images captured on video', *Journal of Experimental Psychology: Applied*, vol.5, pp.339–60.
- Bulthoff, H.H. and Edelman, S. (1992) 'Psychophysical support for a two-dimensional view interpolation theory of object recognition', *Proceedings of the National Academy of Sciences of the USA*, vol.89, pp.60–4.
- Burton, A.M. and Bruce, V. (1993) 'Naming faces and naming names: exploring an interactive activation model of person recognition', *Memory*, vol.1, pp.457–80.
- Burton, A.M., Bruce, V., and Johnston, R.A. (1990) 'Understanding face recognition with an interactive activation model', *British Journal of Psychology*, vol.81, pp.361–80.
- Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B., and Henderson, Z. (2001) 'Human and automatic face recognition: a comparison across image formats', *Vision Research*, vol.41, pp.3185–95.
- Diamond, B.J., Valentine, T., Mayes, A.R., and Sandel, M.E. (1994) 'Evidence of covert recognition in a prosopagnosic patient', *Cortex*, vol.28, pp.77–95.
- Diamond, R. and Carey, S. (1986) 'Why faces are and are not special: an effect of expertise', *Journal of Experimental Psychology: General*, vol.115, pp.107–17.
- Ellis, H.D. and Lewis, M.B. (2001) 'Capgras delusion: a window on face recognition', *Trends in Cognitive Sciences*, vol.5, no.4, pp.149–56.
- Ellis, H.D. and Young, A.W. (1990) 'Accounting for delusional misidentifications', *British Journal of Psychiatry*, vol.157, pp.239–48.
- Gibson, J.J. (1986) *The Ecological Approach to Visual Perception*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Hay, D.C. and Young, A.W. (1982) 'The human face', in Ellis, A.W. (ed.) *Normality and Pathology in Cognitive Functions*, London, Academic Press.
- Hay, D.C., Young, A.W., and Ellis, A.W. (1991) 'Routes through the face recognition system', *Quarterly Journal of Experimental Psychology*, vol.43A, pp.761–91.
- Hirstein, W. and Ramachandran, V.S. (1997) 'Capgras syndrome: a novel probe for understanding the neural representation of identity and familiarity of persons', *Proceedings of the Royal Society, London, Series B*, vol.264, pp.437–44.
- Hole, G. and Bourne, V. (2010) *Face Processing: Psychological, Neuropsychological, and Applied Perspectives*, Oxford, Oxford University Press.
- Humphreys, G.W. and Bruce, V. (1989) *Visual Cognition: Computational, Experimental and Neuropsychological Perspectives*, Hove, Lawrence Erlbaum Associates Ltd.
- Humphreys, G.W. and Riddoch, M.J. (1984) 'Routes to object constancy: implications from neurological impairments of object constancy', *Quarterly Journal of Experimental Psychology*, vol.36A, pp.385–415.
- Johnson, M.H. and Morton, J. (1991) *Biology and Cognitive Development: The Case of Face Recognition*, Oxford, Blackwell.
- Johnston, A., Hill, H., and Carmen, N. (1992) 'Recognizing faces: effects of lighting direction, inversion and brightness reversal', *Perception*, vol.21, pp.365–75.
- Johnston, R.A. and Bruce, V. (1990) 'Lost properties? Retrieval differences between name codes and semantic codes for familiar people', *Psychological Research*, vol.52, pp.62–7.
- Kemp, R., Pike, G., and Brace, N. (2001) 'Video-based identification procedures: combining best practice and practical requirements when designing identification systems', *Psychology, Public Policy and Law*, vol.7, no.4, pp.802–7.
- Kemp, R., Towell, N., and Pike, G. (1997) 'When seeing should not be believing: photographs, credit cards and fraud', *Applied Cognitive Psychology*, vol.11, no.3, pp.211–22.
- Lawson, R. and Humphreys, G.W. (1996) 'View-specificity in object processing: evidence from picture matching', *Journal of Experimental Psychology: Human Perception and Performance*, vol.22, pp.395–416.
- Lederman, S.J. and Klatzky, R.L. (1987) 'Hand movements: a window into haptic object recognition', *Cognitive Psychology*, vol.19, pp.342–8.
- Lederman, S.J. and Klatzky, R.L. (1990) 'Haptic classification of common objects: knowledge-driven exploration', *Cognitive Psychology*, vol.22, pp.421–59.
- Lederman, S.J., Klatzky, R.L., and Pawluk, D.T. (1993) 'Lessons from the study of biological touch for robot haptic sensing', in Nichols, H. (ed.) 'Advanced tactile sensing for robotics', in *World Scientific Series in Robotics and Automated Systems*, vol.5, Singapore, World Scientific Publishing.
- Longmore, C.A., Liu, C.H., and Young, A.W. (2008) 'Learning faces from photographs', *Journal of Experimental Psychology: Human Perception and Performance*, vol.34, pp.77–100.
- Marr, D. (1977) 'Analysis of occluding contour', *Proceedings of the Royal Society of London, Series B*, vol.197, pp.441–75.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company.
- Marr, D. and Nishihara, H.K. (1978) 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proceedings of the Royal Society of London, Series B*, vol.211, pp.151–80.

Megreya, A.M. and Burton, A.M. (2006) 'Unfamiliar faces are not faces: evidence from a matching task', *Memory & Cognition*, vol.34, pp.865–76.

Megreya, A.M. and Burton, A.M. (2007) 'Hits and false positives in face matching: a familiarity based dissociated', *Perception & Psychophysics*, vol.69, pp.1175–84.

Memon, A. and Bruce, V. (1985) 'Context effects in episodic studies of verbal and facial memory: a review', *Current Psychological Research & Reviews*, pp.349–69.

Morrison, D.J., Bruce, V., and Burton, A.M. (2001) 'Understanding provoked overt recognition in prosopagnosia', *Visual Cognition*, vol.8, pp.47–65.

Neisser, U. (1967) *Cognitive Psychology*, New York, Appleton-Century-Crofts.

Pike, G., Brace, N., and Kynan, S. (2001) *The Visual Identification of Suspects: Procedures and Practice: A Publication of the Policing and Reducing Crime Unit*, London, Home Office Research, Development and Statistics Directorate.

Pike, G., Kemp, R., and Brace, N. (2000) 'The psychology of human face recognition', *IEE Electronics and Communications: Visual Biometrics*, 00/018, pp.12/1–12/6.

Robbins, R. and McKone, E. (2007) 'No face-like processing for objects-of-expertise in three behavioural tasks', *Cognition*, vol.103, pp.34–79.

Searcy, J.H. and Bartlett, J.C. (1996) 'Inversion and processing component and spatial-relational information in faces',

Journal of Experimental Psychology: Human Perception and Performance, vol.22, pp.904–15.

Selfridge, O.G. (1959) 'Pandemonium: a paradigm for learning', in *The Mechanisation of Thought Processes*, London, HMSO.

Sergent, J. and Poncet, M. (1990) 'From covert to overt recognition of faces in a prosopagnosic patient', *Brain*, vol.113, pp.989–1004.

Tanaka, J.W. (2001) 'The entry point of face recognition: evidence for face expertise', *Journal of Experimental Psychology: General*, vol.130, pp.534–43.

Tarr, M.J. (1995) 'Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects', *Psychonomic Bulletin and Review*, vol.2, pp.55–82.

Warrington, E.K. and Taylor, A.M. (1978) 'Two categorical stages of object recognition', *Perception*, vol.7, pp.695–705.

Yin, R.K. (1969) 'Looking at upside down faces', *Journal of Experimental Psychology*, vol.81, pp.141–5.

Young, A.W., Hay, D.C., and Ellis, A.W. (1985) 'The faces that launched a thousand slips: everyday difficulties and errors in recognizing people', *British Journal of Psychology*, vol.76, pp.495–523.

Young, A.W., Newcombe, F., De Haan, E.H.F., Small, M., and Hay, D.C. (1993) 'Face perception after brain injury', *Brain*, vol.116, pp.941–59.



Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

PART 2

CONCEPTS AND LANGUAGE

Introduction

5 Concepts

Nick Braisby

6 Language processing

Gareth Gaskell and Helen Brown

7 Language in action

Simon Garrod and Tony Sanford