COGNITION IN ACTION

2nd Edition

Mary M. Smyth
Alan F. Collins
Peter E. Morris
Philip Levy

Cognition in Action

Second Edition

Mary M.Smyth
Alan F.Collins
Peter E.Morris
Philip Levy
Department of Psychology
Lancaster University, UK

Transferred to Digital Printing 2005

Psychology Press Ltd, Publishers 27 Church Road Hove East Sussex, BN3 2FA UK
http://www.psypress.co.uk/
This edition published in the Taylor & Francis e-Library, 2005.

British Library Cataloguing in Publication Data A catalogue record for this book is available from the British Library

ISBN 0-203-01552-5 Master e-book ISBN

ISBN 0-86377-347-8 (hbk)
ISBN 0-86377-348-6 (pbk)

Contents

Introduction

In this book, we approach cognitive psychology by asking what it has to tell us about how people carry out everyday activities. In other words, we ask how do people organise and use their knowledge in order to behave appropriately in the world in which they live. Each chapter in the book starts with an example (which makes up the first part of the chapter title), and then uses the example to introduce some aspects of the overall cognitive system. In this way, the more general psychological functions described in the second part of each chapter title are introduced and explained.

Some of the examples we use are serious ones, like making a medical decision, and others are fairly trivial, like tapping your head and rubbing your stomach at the same time. Some of the examples, like doing mental arithmetic, are simply used to introduce problems and questions about an aspect of cognitive functioning, and are not the topic of the chapter itself. Other examples, like reading a word, do provide the topic for the whole chapter. This is partly because language processing is something we do every day, as well as a central research area for many psychologists. Other everyday activities may not themselves be studied directly by psychologists, or may only be studied as part of a wider enquiry.

Cognition is concerned with knowledge, and cognitive psychology is concerned with the acquisition and use of knowledge, and with the structures and processes which serve this. The cognitive system, although it is complex, normally operates as a whole, and it can be misleading to separate out parts of the system for special attention without emphasising that each part can only be understood properly in its place in the functioning of the whole. Traditionally, textbooks on cognitive psychology have taken topics such as perception, memory and language as major themes, and in doing so have sometimes emphasised the component parts of the system while obscuring its purposes and functions. These divisions also suggest to the reader that cognition presents itself in neat sub-compartments which we can study, rather than that we create the components as we investigate cognition.

Of course, to study any complex system it is necessary to introduce some subdivisions, and the problem is to present these without leading readers to believe that a topic such as memory can be completely understood in isolation. Our solution has been to identify important components in the cognitive system, and to illustrate them through examples of cognition in action. So, for example, all cognition depends on our initial perceptions of the world, and perceptual processing is referred to over and over again throughout the book. However, perception plays an especially important part in reading

and in recognising faces, and by starting with these examples in the first two chapters we are able to highlight the major aspects of the initial perception components of the system.

Unlike other texts, we do not make a rigorous division into "stages" of cognitive processing; rather, we emphasise different aspects of the system in different ways. For example, new information entering the system must be appropriately organised and classified, relevant old information must be retrieved to aid construction and interpretation, the elements of the old and new must be held in a temporary form while the new construction is assembled and decisions made, and a record of what has occurred becomes part of the store of information that is held for future use. Each of these aspects, and many others, can be emphasised in the context of a particular task, so we do not have chapters which treat cognitive processes or stages in isolation from tasks. Thus, while the book moves from what are traditionally "lower" aspects of cognition, such as perceiving faces and producing actions, to "higher" aspects, such as comprehension and problem solving, these are not seen as stages but as important parts of normal human functioning which cognitive psychology has approached in different ways.

Cognitive psychologists have in the past concentrated on experimental laboratory studies. Their research has sometimes moved a long way from the original questions they asked, and a very long way from cognition in action. This can make the topics seem both difficult and irrelevant to the student. Indeed, one of the original reasons for writing a book like this was that our own students found it very difficult to understand why the topics they were studying mattered and how they related to real questions about the mind. Nevertheless, in many cases, we do need to move into the laboratory to control some aspects of normal functioning in order to get a better understanding of how cognitive processes operate. What is important is that such research should give results which can be used to help us answer the original questions, or even to show us that they are not the best questions to ask. We have tried to select such research for this book.

More recently, cognitive psychologists have drawn upon a wider range of sources of evidence. One important source for evidence about the structure of cognition comes from neuropsychology, the study of people who have suffered brain damage. Cognitive models have to be able to account for the patterns of disability which are found following brain damage, and we use this sort of evidence throughout the book. A second major influence on how cognitive psychologists develop and test their theories is the development of computer models, and in particular the development of network models, which have some plausible similarities to the way the nervous system works. These models have changed some aspects of the ways in which cognitive psychologists think about how information is represented, and they also provide a way of testing hypotheses by developing models and comparing the behaviour they produce with the behaviour produced by people in situations similar to those being modelled. Network models appear in the very first chapter of this book, but explaining them and how they are relevant fits into some of the questions raised in Chapter 2, so that is where we have located an introduction to this style of thinking. We also discuss these sorts of models in many of the other chapters.

In writing this book, we have been selective about the research we have used. This means that some topics which are found in other cognition texts do not appear here, and we may have chosen to omit studies which other people feel are important. Our selection depends on the line of argument being made from the examples and questions in real life

to the issues which arise in cognitive research. In some chapters, where there is a long tradition of research that seems to us to have approached problems in intuitively sensible ways, we have reported the work, even if it does not provide complete answers to our questions. This is the second edition of the book, and some chapters and sections of chapters have been extensively rewritten. We have, however, kept to our original intention to discuss relevant work whether it was recent or not, and not to include studies simply because they have been done. Experiments and research programmes have little value by themselves; they only matter when they help us to understand something of how the cognitive system does or doesn't work.

It is easy to take for granted activities such as reading words, recognising people we know and telling tables from chairs. Sometimes it is only when we have had problems with these activities or when we realise that other people have problems with them that we realise that there is something to be explained. In this book, we take the view that there is something to be explained every time anyone reads or fails to read a word, recognises or fails to recognise a face, remembers or fails to remember an intention. We have to ask the question "How did I do that?" when we succeed at these activities in order to understand what it is that cognitive psychologists study. This book does not present the argument that everyday cognition can only be studied in the everyday world, but rather that everyday cognition gives us our problems and our questions. Our answers—however we obtain them—must always relate back to cognition in action.

In writing the second edition, we were unable to call on Andy Ellis who was one of the authors of the first edition. His ideas, his examples and even some of his very words remain in this version. It wouldn't be the same without him. The errors, as always, remain our own.

1
Recognising Faces: Perceiving and Identifying Objects

It is a busy Saturday afternoon in your town. The streets are swarming with shoppers pushing and shoving. You are trying to find a pair of shoes you like and wondering why on earth you didn't do your shopping midweek when things were quieter. In the distance, you notice two people walking towards you. The one on the left you recognise immediately as your grandmother; the one on the right you do not recognise.

What could be a more commonplace and everyday occurrence than recognising the face of someone you know? We do it all the time—at home, at work, watching television, in town. "But", asks the cognitive psychologist, "*How* do we do it? *How* do we recognise the lady approaching us as granny? What processes go on in our minds that allow us to identify the lady on the left as familiar while rejecting the person on the right as unfamiliar?" Person and face recognition must be a matter of achieving a match between a perceived stimulus pattern and a stored representation. When you get to know someone, you must establish in memory some form of representation or description of his or her appearance. Recognising the person on subsequent occasions requires the perceived face to make contact with the stored information, otherwise the face will seem unfamiliar.

As cognitive psychologists try to develop a more specific account of recognising a person, problems start to arise. What form does the internal representation of a familiar face take? How are seen faces matched against stored representations? How does seeing a familiar face trigger the wider knowledge you have about that person, including his or her name? When you see a familiar person, he or she is often moving against a complex visual background: How does your visual system isolate elements of the whole visual scene as constituting one object (a person) moving in a particular way at a particular speed? These are some of the questions a complete theory of visual processing should be able to answer, and some of the questions that are addressed in this chapter. We use face recognition in this chapter to introduce some of the important questions about how we perceive and recognise objects in general, not just faces. We will consider both the general issue of how we recognise objects, including people, and the more specific questions concerning how we recognise people we know.

RECOGNISING FAMILIAR PATTERNS

There are lots of faces visible in the shopping crowd in our example, but you recognise only one of them. How? A ploy commonly adopted by cognitive psychologists when trying to understand how the mind performs a particular task is to ask how we could create an artificial device capable of performing the same task. How could we, for example, program a computer to recognise a set of faces and reject others?

First of all, the computer would need to somehow memorise the set of faces it had to recognise. It would then need to compare each face it saw (through a camera input, for example) with the set stored in memory to see if there was a match. If a satisfactory match was achieved, the face would be "recognised"; if not, it would be rejected as unfamiliar.

Now, within those broad outlines, there are a number of options available regarding the possible nature of the representation of each face to be stored in memory and the manner in which the perceived face could be compared against the stored set. The stored set of faces could form a set of templates, with the new face being matched to each template and recognition occurring when a complete or nearly complete match of template and pattern took place. Perhaps recognising a particular face would require the stimulation of a particular pattern of cells within the retina of the eye. Different patterns of stimulation would be stored for each known face. Pattern recognition systems along these lines have been used for many years in, for example, the mechanical reading of the numbers upon bank cheques.

Template mechanisms of pattern recognition are relatively simple to set up. However, they have serious limitations which suggest that they are not the mechanism used by the human perceptual system when recognising familiar objects. Problems arise as soon as there is any change in the original stimulus. For example, if you see your grandmother from a different distance to that for which the template was set up, then a smaller or larger image will be projected onto the retina of your eye and will stimulate a different set of cells. Similar problems arise if you see your grandmother from an angle different to that for which the template was created. Also, people change in their appearance—their hairstyles, their spectacles, their faces age, and so on. While such changes can cause us problems in recognition, we do normally still identify our acquaintances. However, the mismatch with any template would be sufficient for it to fail to be selected.

Template-based systems of pattern recognition can be elaborated. They can, for example, include more than one template, so that common views of the same object can be recognised. Face recognition might include templates for views of the full face, three-quarter (portrait) and profile angles. There is evidence that cells in the brains of monkeys are differen-tially sensitive to such alternative views (Perrett et al., 1984). It is possible to "normalise" a new pattern until it is a standard size and orientation before it is matched to the templates. Some theorists of face recognition have considered template accounts (e.g. Ellis, 1975). However, most researchers have looked to more sophisticated ways in which the information about known objects and people might be matched to a newly experienced pattern.

One alternative might be the storing of the description of a person's face in terms of a list of *features* (a feature being a property of an object that helps discriminate it from

other objects). Granny's face might then be held in memory as a feature list something like:

+ white hair
+ curly hair
+ round face
+ hooked nose
+ thin lips
+ blue eyes
+ round gold-rimmed spectacles
+ wrinkles

and so on. The features of each face to appear before the camera could then be compared against the stored list. If all features agreed, then the face would be recognised as granny's, but if the person before the camera had, say, a long, thin face rather than a round one, it would be rejected as unfamiliar. This would be a *feature-based* model of recognition. One of the advantages of such models is that they do not tightly specify how the features go together as is the case with template models. For example, the same set of features can be recognised from many different views of the same face.

There is no denying that features play a role in face recognition, or that some features are more important than others. In free descriptions of unfamiliar faces, subjects utilise features, mentioning the hair most often, followed by eyes, nose, mouth, eyebrows, chin, and forehead in that order (Shepherd, Davies, & Ellis, 1981). As faces become familiar, there is evidence of a decreasing reliance on external features such as hairstyle, colour and face shape towards a reliance on the internal features of eyes, nose and mouth (Ellis, Shepherd, & Davies, 1979). This may be because hairstyle in particular can change, and so is a relatively unreliable cue to recognition, whereas internal features are comparatively stable and reliable.

The problem with any simple feature-based model is illustrated by Fig. 1.1. A "scrambled" face contains the same features as a normal face, but their *configuration* has been altered. Although it may be possible to recognise the scrambled face of a well-known person, it is much harder than recognising a normal face. Also, as Fig. 1.2 shows, varying the configuration of a fixed set of features can substantially alter the appearance of the face. So, while template models may be too specific about the details of each face, a simple feature list model would not be specific enough. A satisfactory model of face recognition will take into account the configuration of the features but will be sufficiently flexible that it can recognise the same face despite the patterns actually experienced varying considerably, because the person is being seen from one of many possible angles or distances.

FIG. 1.1 The scrambled face of a well-known person illustrates the importance of configuration in pattern recognition. Reproduced with permission from Bruce and Valentine (1985).

You will be beginning to appreciate the challenge faced by object and person recognition systems, whether our own or those that might be created, for example, to allow robots to behave like humans as they do in science fiction stories. Such recognition in a world of moving, changing objects is very difficult. However, the challenge is greater still than we have discussed so far. How do we even manage to perceive an object as an object? We will consider this even more basic question in the next section.

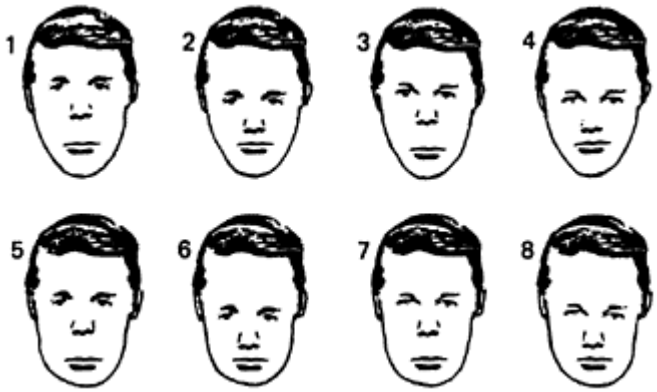

FIG. 1.2 Each pair of faces (1 and 2; 3 and 4; 5 and 6; 7 and 8) differ only in the configuration of their internal features. Adapted with permission from Sergent (1984).

SEEING OBJECTS

Your grandmother is moving through a crowd of shoppers carrying a couple of bags. Parts of her periodically disappear from view when she passes behind a bench seat, or when another shopper passes in front of her. Your visual system is confronted with a kaleidoscope of patches of light of different colours, reflecting off surfaces of objects at varying distances, moving in varying directions at varying speeds. What you *perceive,* however, is a coherent scene composed of distinct objects set against a stable background. This unified impression is the *end-product* of processes of visual perception which psychologists have sought to understand.

Some of the first psychologists to be interested in how we perceive one part of a visual display as belonging with another were the German Gestalt psychologists. From 1912 onwards, Gestalt psychologists, led by Wertheimer and his students Kohler and Koffka, concentrated upon the way in which the world we perceive is almost always organised as whole objects set against a fixed background. Even three dots on a page (see Fig. 1.3) will cohere as a triangle. Our perceptual systems are organised to derive forms and relationships from even the simplest of inputs. The Gestalt psychologists argued that our perceptual systems have evolved to make object perception possible. They set out to describe the principles that the perceptual system uses to group together the elements in the perceptual field. Subsequently, those attempting to model object perception (e.g. Marr, 1982) have incorporated these principles into their models, as we describe later in the chapter.

FIG. 1.3. Three simple dots on a page cannot but be seen as a triangle.

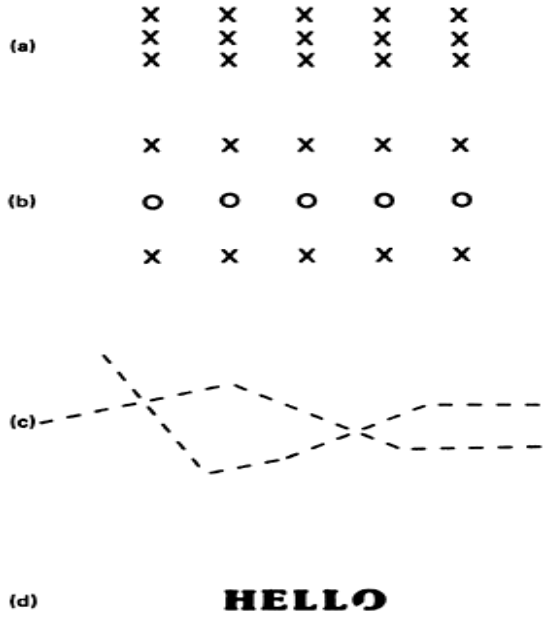

FIG. 1.4 Examples of Gestalt principles in action. (a) *Proximity:* the arrangement of the crosses causes them to be perceived as being in columns rather than rows. (b) *Similarity:* the similarity of the elements causes them to be perceived as being in rows rather than columns. (c) *Good continuity:* causes you to interpret this as two continuous intersecting lines. (d) *Closure:* the gap in the "O" is perceptually completed.

The Gestalt psychologists formulated several principles to describe the way in which parts of a given display will be grouped together (see Fig. 1.4a). However, grouping is modified by the similarity of components. So, in Fig. 1.4b, the noughts and crosses tend to be seen in lines because they are similar. In Fig. 1.4c, the lines of dashes are seen as crossing one another, rather than meeting at a point turning at an angle and moving away. This illustrates the Gestalt principle of *good continuation,* which maintains that elements will be perceived together where they maintain a smooth flow rather than changing abruptly. In a similar way, the perceptual system will opt for an interpretation that produces a closed, complete figure rather than one with missing elements. Sometimes this can lead to the overlooking of missing parts in a familiar object. If you had not been primed to look for it by the text, it would be easy to conceptually complete the word and overlook the gap in one of the letters in Fig. 1.4d. Other perceptual preferences highlighted by the Gestalt psychologists were for bilaterally symmetrical shapes (e.g. Fig. 1.5a). Other things being equal, the smaller of two areas will be seen in the background, and this is enhanced by them being in a vertical or horizontal arrangement (see the black and white crosses in Figs 1.5b and c).

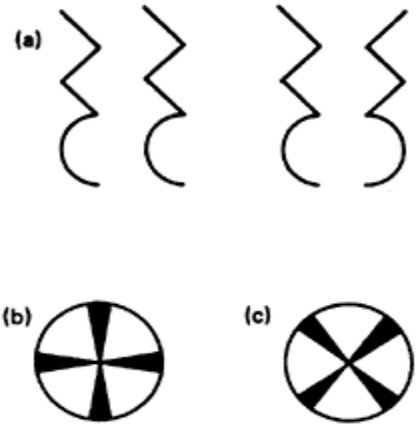

FIG. 1.5 (a) Organisation by lateral symmetry. The symmetrical form on the right is much more easily perceived as a coherent whole than the asymmetrical form on the left. (b) The preference here is to perceive the smaller area as the figure and the larger area as the ground, i.e. as a black cross on a white background. (c) If the larger areas is to be perceived as the figure

(i.e. a white cross on a black background), then orienting the white area around the horizontal and vertical axes makes this easier.

To summarise their principles, Wertheimer proposed the Law of Pragnanz. This states that, of the many geometrically possible organisations that might be perceived from a given pattern of optic stimulation, the one that will be perceived is that possessing the best, simplest and most stable shape. Sometimes, the input can be interpreted in more than one way and the result is a dramatic alternating in our perception. The face-vase illusion (Fig. 1.6) devised by the Gestalt psychologist Rubin is a well-known example. The information in the picture allows it to be interpreted either as a vase or as two faces. When the interpretation shifts, the part that had been the figure becomes the background, and vice versa.

Although we have illustrated the Gestalt principles using very simple examples and illustrations, their application to normal, intricate visual processing is in the way they assist the visual system to unite those components of the visual array that constitute single objects. There are other cues that assist in this unification. All the cues discussed so far apply to stationary objects, but additional cues arise when an object moves. If an object is moving, it will cover progressively more of the visual field if it is approaching, less if it is going away, and will successively obscure and reveal the background over which it passes. This movement is a major source of object perception. Elements that move together are usually perceived as being part of the same object, a principle known to the Gestalt psychologists as *common fate*.

FIG. 1.6 Rubin's well-known ambiguous figure, which can be seen as either a black vase against a white background, or as two white faces against a black background (but not both at once).

How do all these principles apply to our example of recognising our grandmother? To see her as an object at all we must perceive which parts of our current visual array are a part of her body and which are not. Granny's good continuation, the common fate of her parts and so on, will all help you to unify those components of your current perceptual field which belong together as the parts of the single object that is your grandmother. It is convenient for the purpose of the illustration, earlier, to isolate each cue and demonstrate it in simplified form, but in the real world these cues are all operating together, and their function is to assist the visual system in identifying objects in the visual scene with a view to recognising them for what they are. So far, we have been considering how aspects of the visual array can be used, as summarised in the Gestalt principles, to identify objects from their backgrounds. Such identification is often not in the interest of a carnivorous animal seeking its prey, or of the prey itself. The evolution of camouflage in the natural world can be analysed as a variety of attempts to disrupt and confound the Gestalt principles, making an animal hard to distinguish from the background scene.

DISTANCE AND MOVEMENT

We have considered some of the ways in which an object may be separated from its background, but much more needs to be perceived than that about the behaviour of animate or inanimate objects if the perceiver is to survive long! Two very important properties of the object are its distance and the way it is moving. As soon as you become aware of granny's presence in the crowd, you also have an impression of how far away she is. Judgement of the relative distances of visible objects is an important aspect of perception.

Binocular Depth Cues

One source of distance information is the stereoscopic information provided by the two slightly different views of the same scene obtained by our two eyes. Given the disparity between the images to the two eyes, it is possible to calculate the distance of an object, because the closer an object is, the greater is the disparity between the two different views of it.

Stereoscopes were invented in the 1830s by Wheatstone and have been popular at various times since for the vivid three-dimensional experience they produce. A different picture is presented to each eye, each picture representing what would be seen if the actual objects were present in three dimensions. So, if the pictures are photographs, the two photos are taken from positions a few inches apart, thus reproducing the views from our two eyes.

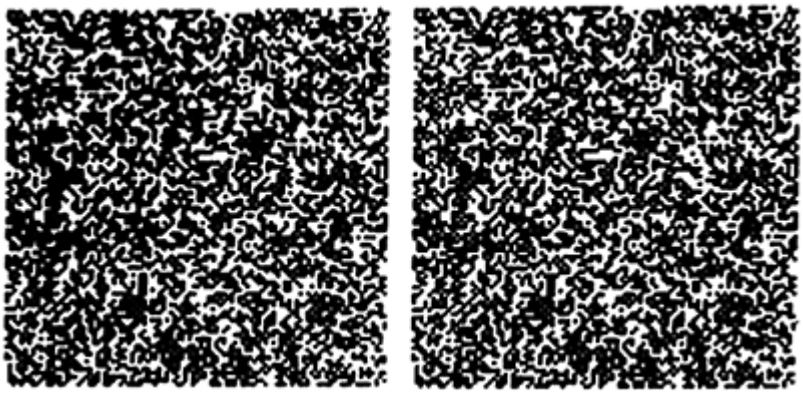

FIG. 1.7 Julesz dot patterns (see text for explanation). Reproduced with permission from Julesz (1964).

How exactly is the discrepant information from the two eyes combined to allow depth to be computed? The traditional view was that the images from each eye were separately processed, identifying the objects in the scene, and only then fused together (e.g. Sherrington, 1906). However, the work of Julesz (1971) has suggested that this is not so. Julesz developed the *random-dot stereogram* (see Fig. 1.7). This consists of patterns of black and white dots. Viewed without a stereoscope, the two patterns shown in Fig. 1.7 look similar. If, however, the patterns are projected one to each eye, a central square of dots will be seen floating closer to the observer. The reason is that the two random-dot stereograms are not identical. The right-hand one has a square part of the left-hand pattern shifted to the right and the space remaining filled with random dots. To the retina of each eye, this provides the same information that would be seen if that square was

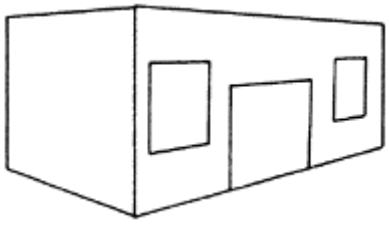

FIG. 1.8 Converging lines give a strong impression of depth.

Monocular Depth Cues

actually in front of the rest of the main square, and it is seen in that way when the stereoscopic information is combined. In fact, it is only possible to see the square

stereoscopically. What Julesz random-dot stereograms demonstrate is that one can see stereoptically without having to first recognise and fuse an object separately for each eye. When seen separately we cannot see the object, the square, which only appears when the images are combined.

We have discussed depth perception using two eyes, but even with one eye closed we can normally judge how far away an object is. With one eye closed we could still shake hands with our grandmother's friend! There are many cues to distance in most static visual scenes. First, there is the relative size of known objects—the farther away your grandmother is, the smaller the area of the retina upon which her image is projected. Second, things that are closer will often be superimposed upon and obscure parts of the view of things farther away. Shadows give impressions of solidity and depth to individual objects. The texture of the things we perceive becomes less obvious and finer at greater distances. For example, in a field, we can see the details of the blades of grass near our feet, but such detail is lost as we look farther away. This texture gradient is a strong cue to distance. Especially in a world of rooms and buildings with their straight lines, right angles and flat planes, perspective is another strong depth cue. As they recede into the distance, parallel lines converge, and line drawings that incorporate such features give a convincing impression of depth (see Fig. 1.8). Furthermore, the intersection of edges and the obscuring of parts of objects provide further cues to depth. Where one edge meets another in a T-junction, one object is normally in front of another. As we discuss later, this is a feature that has been used in scene analyses.

By careful building, it is possible to construct visual displays that set monocular cues against one another and produce visual illusions. A famous one is Ames' room (Ittelson, 1952). In Fig. 1.9, photographed inside the room, it appears as if the person on the right is very much bigger than the one on the left, yet both are actually normal-sized adults. In Ames' room, the normal depth cue of our knowledge of relative sizes is overwhelmed by the manipulation of the perspective provided by the decorations on the walls. To the viewer, the room looks rectangular because the decorations resemble doors and windows as they would appear in a normal rectangular room. In fact, the room increases in distance and height away to the left and the decorations are trapezoidal, not rectangular. The person who appears smaller is much farther away than the other one. Ames' room illustrates how, when depth cues conflict, our perceptual system will sacrifice and re-scale familiar features to produce a consistent representation.

A moving person or object supplies additional information to help in the judgement of distance. The farther away a moving object is, the slower it will move through the visual field. When we are ourselves moving, then the world that we see seems to flow past us. The speed and direction of the optic flow provide excellent information about the direction and speed of our travel. This optic flow is illustrated in Fig. 1.10. Gibson (1966; 1979) has argued strongly that too much emphasis in psychology has, in the past, been placed on the perception of static, very simple displays by static observers in a highly uniform, often visually degraded environment. Our perceptual systems have actually evolved to cope with a visually extremely rich world in which we are constantly moving and experiencing changes in the visual array, and the cues afforded by movement and change are important in determining the interpretation we place upon a visual scene.

FIG.1.9 Ames' room (see text for explanation). Reproduced with the permission of Eastern Counties Newspapers Limited.

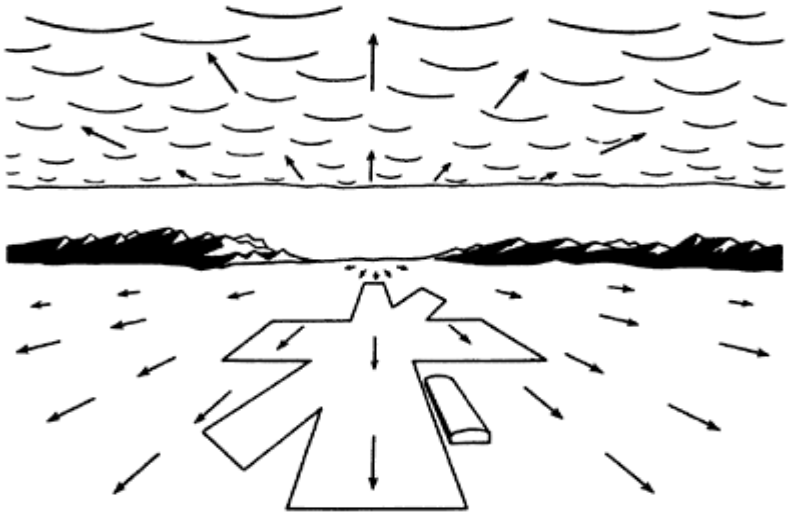

FIG. 1.10 The optic flow field for a pilot landing an aeroplane. Reproduced with permission from Gibson (1950).

When it comes to grouping elements of an array together as components of a single object, or judging how far away that object is, then the processes involved are likely to be the same whether or not you recognise the individual object concerned (they will treat granny and the unfamiliar lady walking alongside her alike). For the remainder of this chapter, we return to the processing of familiar objects, concerning ourselves with such things as how, having recognised an object, you retrieve the relevant information about it that you have stored in memory (including its name), and what role context plays in the recognition of familiar objects.

IDENTIFYING OBJECTS

Scene Analysis Programs

In the 1960s and 1970s, researchers in artificial intelligence tried to write computer programs that would be capable of identifying objects. Such attempts at the computer simulation of human skills are often stimulated by both practical and theoretical motives. The practical one is the need for any mechanical device that moves around and handles newly encountered objects to be able to perceive those objects. The theoretical aspect is that the need to produce a working computer simulation forces researchers to consider all the elements of the problem, including some which will not have occurred to the armchair speculator on the problem. To simplify this problem, they concentrated upon recognising evenly lit, smooth-sided blocks and prisms. Guzman (1968), Clowes (1971) and Waltz (1975) developed programs which analysed the lines that were present in, for example, the display in Fig. 1.11.

The analyses concentrated upon line junctions, with different types of junctions being classified as indicating different relationships between the blocks. So, an arrow junction (A in Fig. 1.11) generally involves planes from the same body, whereas T-junctions (B in Fig. 1.11) normally occur where the crossbar and shaft of the T are part of different bodies. Edges were labelled as *lower* (pointing outwards), *concave* (pointing inwards) and *occluding* (i.e. occluding other bodies by being the outer edge of the solid as seen by the observer). In this way, impossible figures such as that in Fig. 1.12 could be rejected by the programs.

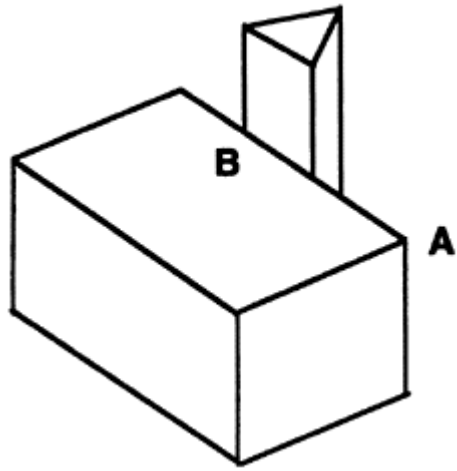

FIG. 1.11 A block and prism suitable for analysis by scene analysis programs. A is an arrow junction and B is a T junction.

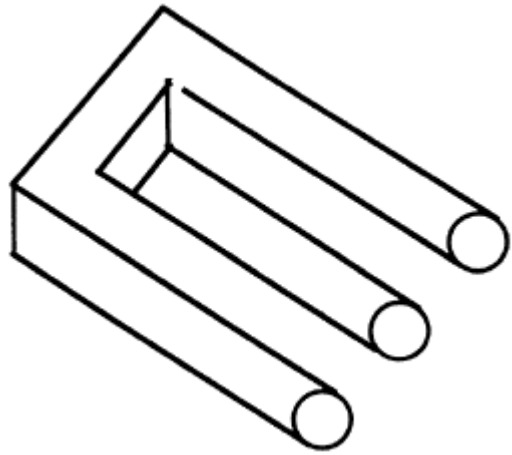

FIG. 1.12 An impossible figure that would be rejected by scene analysis programs.

While interesting for their analysis of the relationship between edges in solid bodies, these programs were limited by their artificial world. Real scenes rarely have straight lines with angular junctions and are much more misleading with their textures, shadings and internal details of objects. There will always be limitations on artificial worlds, and to cope with the complexity, variability and richness of the real perceptual world a more

sophisticated analysis was required. One such analysis was provided by Marr and his associates (Marr, 1982).

Marr's Theory of Vision

In this section, we will introduce the very influential computational theory of vision proposed by Marr and his colleagues. Marr's work was summarised in his book *Vision,* published posthumously in 1982 following his death from leukaemia at the age of 35.

Marr's work is important for several reasons. Not only did he provide a theoretical account of the visual processes, he also highlighted more general issues concerning the types of explanation that we should be seeking. Unlike earlier research, he started from the question of what a *general* theory of vision would require. The fundamental question that he asked is the central mystery of visual perception. How can our processing system take the patterns of light intensity stimulating the retinas of our eyes and from them derive the representations of a world made up of three-dimensional objects that is the form of our conscious perception of the world? Much of Marr's work was directed towards answering this question. However, he realised that the form that an answer must take depends upon asking the right questions and seeking an appropriate level of explanation. This is an insight applicable to all of cognition, not just to vision.

What Does an Explanation of Vision Involve? Marr asked what was the purpose of vision? This may seem too obvious to need attention, but the visual systems of animals such as frogs or insects such as flies have evolved to be integrated into the basic needs of catching prey, escaping danger, etc., rather than to provide a passive projection of some external world to be contemplated in repose by the animal. Our visual systems have evolved to selectively represent certain aspects of our worlds and not others. In Marr's words, "vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information" (Marr, 1982, p. 31).

But what would constitute a description of the visual process? Would it be a description of the interconnections of the neurons of the brain that are involved in vision? That would certainly form part of a complete understanding of the visual system. But would it be sufficient? Marr asked, suppose that one actually found the apocryphal grandmother cell, a cell that fires only when one's grandmother comes into view, would that really tell us anything much at all? It would not tell us *why* such a cell existed in the system, nor how it used the outputs of other cells to create its unique property of recognising grandmother. There is, therefore, much more to understanding vision than this. Marr commented that trying to understand perception by studying only neurons was like trying to understand bird flight by studying only feathers. The structure of feathers and birds' wings make sense only if we understand aerodynamics. Similarly, we will understand the interconnections and firing of the cells of the nervous system only if we can place it in the context of the functions that it is serving. So, an understanding of vision or any cognitive process will require an understanding of the functions served with the life of the individual.

Levels of Explanation. When we come to analyse any information-processing system, Marr argued that we need to recognise that there are at least three levels of explanation that need to be considered. One is the physical mechanism itself, what Marr, as a

computer scientist, called the *hardware implementation*. For the visual system, this is the eye and the cells of the brain that process the output from the eye. However, the activities of these cells can only be understood if we know what is the *goal* of the processing—what has the system evolved or been designed to achieve? For visual processing, Marr identified the underlying task as to reliably derive properties of the world from images of it. When fully specified, this is what Marr called the *computational theory* of vision, and much of his book entitled *Vision* (1982) was directed towards specifying this theory. Marr identified a third level at which an information-processing device needs to be understood. This is the way in which the input and output of the system are represented and the *algorithm* that accomplishes the transformation.

We can take Marr's example of a cash register in a supermarket to help to explain these three levels. The hardware implementation of a cash register has changed over the years. Mechanical machines have been replaced by increasingly sophisticated electrical machines. So, the hardware implementation of a cash register can take many forms, and its implementation in a modern supermarket is different from that in a store 20 years ago. However, the reasons for the cash register's processing and the computational theory underlying its design remain the same. The register carries out addition—that is its computational theory. However, the way in which it actually calculates the addition may vary from machine to machine, because there are several ways to represent the values—for example, all calculations could take place in the decimal system. That is the form in which the customer understands the prices and will expect to be told the total. However, decimal codes may be converted to binary codes to suit the computing hardware. Bar codes provide another alternative input requiring conversion to a price. What about the computing algorithm? A common algorithm for addition is to combine the least significant figures first, then the next, working from right to left and "carrying" if the sum exceeds 9 in decimal, or 1 in binary. This same algorithm will be used by many very different processing systems—a person carrying out mental arithmetic, an old mechanical calculator or a modern electronic calculator. On the other hand, different algorithms might be used to reach the same result—so long as they implement the theory of arithmetical addition.

To summarise, there are the three levels of (1) computational theory, (2) realisation through representation and algorithm and (3) the hardware implementation. They are interconnected, since the computational theory restricts the means of realisation, while the algorithms and representations restrict the possible hardware implementations. But all three levels need investigation. Cognitive psychologists concern themselves with the first and second levels, while physiologists and neuroanatomists try to map out the biological hardware of cognition.

The Computational Theory Underlying Vision. Marr argued that we should ask what is the computational theory underlying the visual system? How are the computations carried out and what physiological and biochemical processes allow this to happen? Marr recognised that perceiving objects was a major purpose of the system. This might seem devious, but it is not necessarily so. The visual systems of some more primitive animals seem to be particularly sensitive to movement, as if that was their main purpose. As we showed earlier, the human visual system, as the Gestalt psychologists pointed out, seems to seek for objects within the visual array. Marr argued that what was required was "a theory in which the main job of vision was to derive a representation of shape". Other

aspects of vision (colour, texture, etc.) he saw as secondary. Marr suggested that the processing would be *modular,* with parts functioning as independently as possible to minimise the problems with failures in any module.

Marr now faced the problem of how stimulation of the cells of the retina by light leads to the perception of objects. At the retina, cells called *rods* and *cones* fire if stimulated by light. These cells interconnect with others, still within the retina, which they excite or inhibit depending upon the pattern of the light stimulation. These cells, in turn, interconnect, exciting and inhibiting cells further back in the visual pathway to the visual cortex.

How could all this lead to our conscious perception of objects? Not simply! However, Marr proposed that there are several stages in this processing. At each stage, a representation is constructed as the result of that stage's processing. Algorithms operate upon the representation derived from the previous stage to produce a new representation. These successive processes gradually transform the information from the pattern of intensity of light stimulating the retina into three-dimensional representations of objects. At each stage, the processing makes use of properties of the representation and consistencies in the world.

The starting point is the retinal image. It gives the distribution of the intensity of stimulation across the retina. From the retinal image, information about the organisation and relationships of *changes* in intensity are, according to Marr, explicitly extracted. This leads to the first stage, which Marr called the *primal sketch*. The type of information that it contains makes possible the detecting of surfaces, as we will describe shortly. The next stage in Marr's theory he called *the* $2\frac{1}{2}$*dimensional* $(2\frac{1}{2}\text{-}D)$*sketch*. In this, the orientation and rough depth of visible surfaces emerges—a "picture" of the world, but only from the perceiver's viewpoint. In the final stage, the *three-dimensional model* representation of the shapes and their relationships form a model of the external world. This model is independent of the particular orientation of the original stimulation of the retina so that, for example, familiar objects will be recognised irrespective of the particular angle from which they are seen.

The Primal Sketch

The primal sketch is made up of a very large number of what Marr called *primatives*. These are derived from the retinal image through computational transformations. These primatives indicate edges, bars, terminations of those edges, blobs, etc. Then, further processing can be carried out on these primatives. For example, where adjacent primatives have a common property (e.g. the same orientation), they are replaced by *tokens* to represent this common property. These then form boundaries between the parts of the *full primal sketch*.

To illustrate the extraction of primatives, we will consider how the system locates what Marr called *zero crossings,* which are of particular importance in his theory, since they are the basis of the *raw primal sketch* prior to the assignment of *tokens* in deriving the full primal sketch.

Zero crossings indicate the sudden change in the intensity of stimulation of the retina. They are used in Marr's model in the identification of edges. When identifying objects, locating their edges is, clearly, especially important. It is common for the intensity of

stimulation to change suddenly at an edge, rather than gradually as it does across a continuous surface.

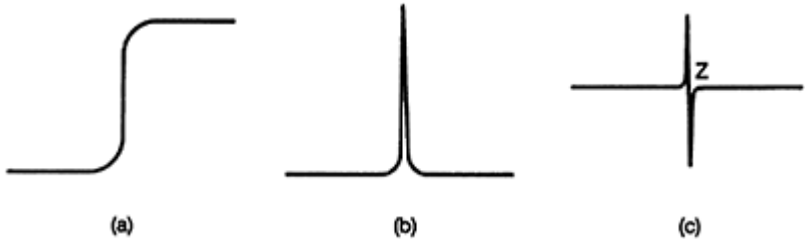

FIG. 1.13 (a) Change in intensity of stimulation at an edge. (b) The rate of change of that stimulation. (c) Rate of change of (b), with a zero crossing at z.

Figure 1.13a represents how the intensity of stimulation on a line from A to B across a small part of the retina might change from low stimulation to high stimulation where there is the edge of an object. It is possible to examine this change in stimulation in a number of ways. If, instead of plotting the intensity of stimulation from A to B, we plotted the *rate of change* in intensity—that is, how quickly it was increasing or decreasing—that would show a sharp rise where the intensity began to increase and a sharp fall when it levelled off. This is shown in Fig. 1.13b. If we now consider further the rate of change shown in Fig. 1.13b, it rapidly increases to a positive peak, then rapidly decreases to a negative peak and then returns to level. This is illustrated in Fig. 1.13c. In moving from its positive to its negative peaks, the graph displays a *zero crossing,* i.e. its value goes through zero. This is what Marr means by zero crossing. Essentially, it is a place where the intensity of stimulation of the retina changes abruptly. It is these crossings that are identified and their position recorded in the primal sketch. By so doing, Marr showed that the outline of shapes can be abstracted from the retinal image.

The identification of zero crossings may sound complicated, but it is easy to calculate mathematically. More importantly, it is possible to use actual filters operating in this way to analyse real photographs and identify lines, edges, etc., from them, so demonstrating that the theory is practicable. It is also possible to propose ways in which the cells that receive signals from the retina may identify zero crossings.

Marr (1982; see also Marr & Hildreth, 1980) analysed images with a range of spatial filters, sensitive to differing spatial frequencies. Figure 1.14 illustrates the result of applying two spatial filters to the same image. In Marr's approach, several spatial frequencies were compared and used to confirm the existence of edge features. If zero crossings were found in the same position for the outputs from a number of spatial frequencies, this was evidence that the zero crossings were the result of edges in the external world.

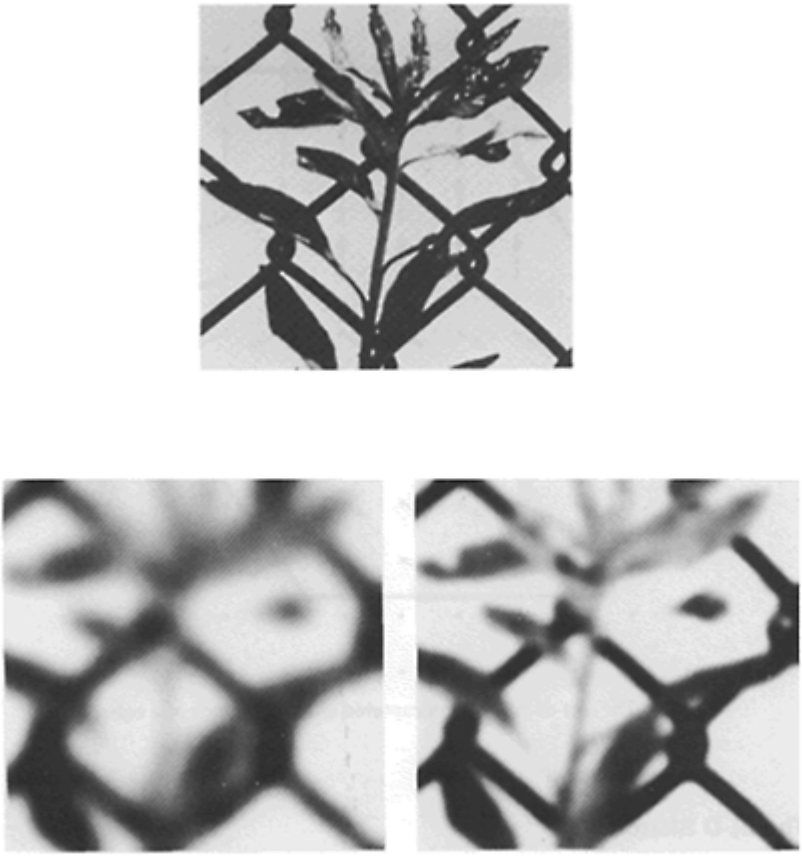

FIG. 1.14 An image (above) blurred by Gaussian filters of two different widths (below). The more blurred picture is produced by the wider filter. Reproduced with permission from Marr and Hildreth (1980).

Much remains to be clarified and elaborated in the underlying procedures that lead to Marr's primal sketch. Georgeson and Shackleton (1989), for example, have shown that the spatial filters need to be more sophisticated than earlier workers assumed. However, the approach has proved very promising. For example, Marr and Poggio (1976) were able to propose a solution to the problem of how stereoscopic depth could be seen using Julesz random-dot patterns. Their algorithm has been implemented as a connectionist program that successfully identifies the stereoscopic pattern (see Chapter 2 for an account of connectionist models).

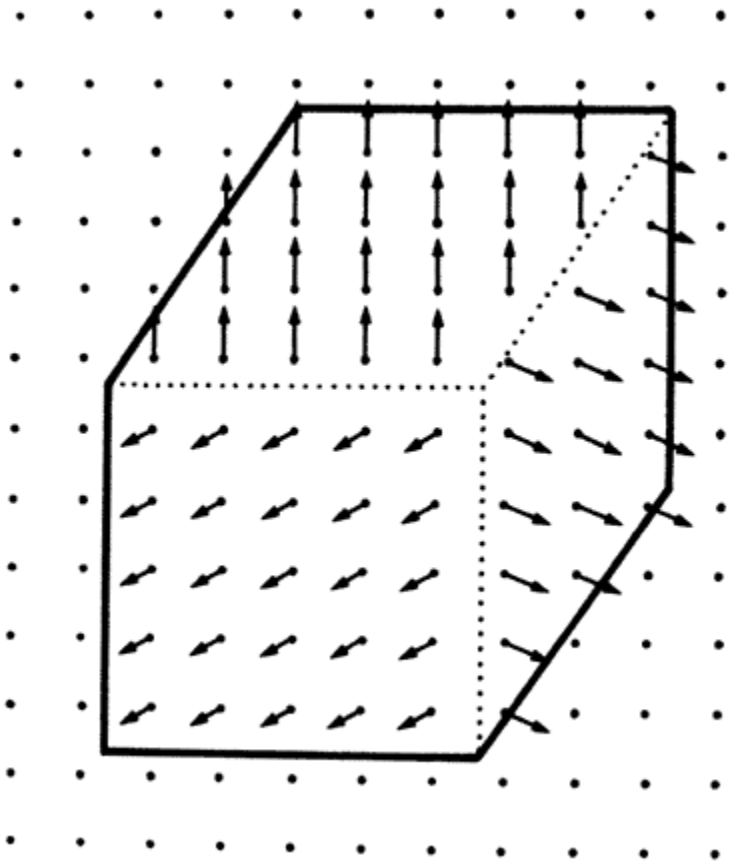

FIG. 1.15 The sort of information captured in a $2\frac{1}{2}$-D sketch of a cube.

The $2\frac{1}{2}$-D Sketch

From the representation provided by the full primal sketch, Marr hypothesised that a description of the visible surfaces in the environment, from the viewpoint of the observer, could be produced. This takes place by applying further algorithms to analyse the patterns of tokens in the primal sketch. The resulting representation, which he called the $2\frac{1}{2}$-D sketch, captures details of orientation and relative depths, but only local changes in depth are represented accurately. Figure 1.15 illustrates the sort of information captured in the $2\frac{1}{2}$-D sketch.

The representation in the $2\frac{1}{2}$-D sketch has new primatives representing orientation that Marr depicts as “needles” with their length and angle representing the degree of tilt and the direction in which the surface slants, respectively. The sketch is called $2\frac{1}{2}$-D because

it contains some, but not all, depth information. Marr still called it a "sketch" because it was from the observer's viewpoint.

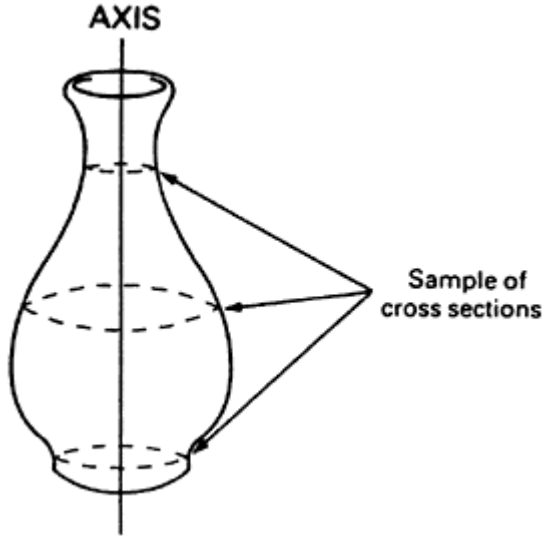

FIG. 1.16 One example of a generalised cone. The shape is created by moving a cross-section of constant shape but variable size along an axis.

Marr and Nishihara's Object Recognition Theory

So far, the analysis from the retinal image to the $2\frac{1}{2}$-D sketch is one of the perception of a scene. Within that scene, however, there are objects that we can recognise despite considerable variation in their orientation. You recognise your grandmother whichever way she is facing, or if we meet her on a staircase and are looking down or up at her.

Marr and Nishihara (1978) based their theory of object recognition upon the assumption that many shapes can be described as *generalised cones*. A generalised cone has an axis along which a shape is moved to map out the contours of the object. The shape must remain the same (e.g. a circle), but its size can vary; so, for example, the base in Fig. 1.16 is a generalised cone. Marr and Nishihara assume that more complex shapes can be accommodated as a hierarchy of 3-D models, each with its own generalised cone around a specific axis. The analysis for a human shape is shown in Fig. 1.17. These hierarchies, called 3-D model descriptions, allow general distinctions between, for example, types of animals to be made.

Marr suggested that the contours of shapes can be derived from the $2\frac{1}{2}$-D sketch and that these can be analysed into generalised cones. This could be used to access stored

information about 3-D model descriptions of different types of objects, and these, in turn, can guide the clarification of the analysis of the object being perceived.

Despite progress in understanding the nature of the objects we perceive, there is very much more to be discovered about how objects as complex as those with which we habitually and effortlessly live are identified by our perceptual systems. The theories of Marr deal with only the most general

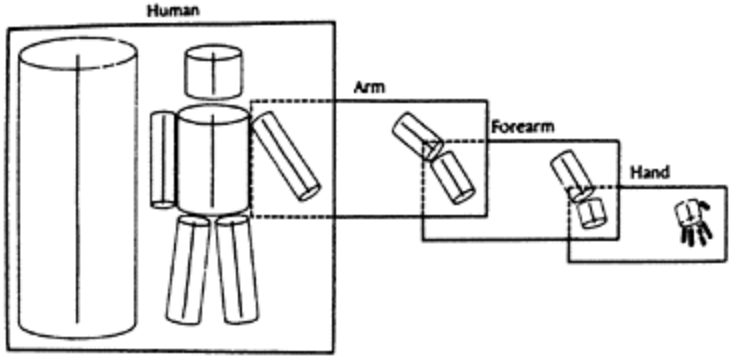

FIG. 1.17 A hierarchy of 3-D models. Each box shows the major axis for the figure of interest on the left, and its component axes on the right. Reproduced with permission from Marr and Nishihara (1978).

classification of objects, while as we know we can recognise the most subtle differences between objects or faces. Nevertheless, as we discuss in the next section, Marr's accounts of the stages of representations have influenced some theories of face recognition.

RECOGNISING AND NAMING FAMILIAR FACES

We have spent some time considering how the visual system may identify objects. Now we want to turn to research that has taken the specific task of recognising that a special sort of object, a human face, is a face that has been seen before. Accompanying this recognition is often the recall of much that you know about the person. When you see granny walking towards you, the act of recognition is typically accompanied by more than just a feeling of familiarity. Something of what you know about her also springs to mind. You may remember that she is a little hard-of-hearing and resolve to speak clearly; you may remember that she is critical of what she considers untidy appearance and surreptitiously try to spruce yourself up; you may remember that she holds strong political opinions somewhat different from your own and make a mental note to steer clear of contentious topics. You will also be able to remember her name.

Errors in Person Recognition

This is what *normally* happens, and what *should* happen, but we all know that our cognitive processes sometimes let us down. Young, Hay and Ellis (1985) persuaded 22 volunteers to keep diaries of their everyday errors and problems in person recognition. A total of 1008 incidents were recorded. They fell into several different categories, four of which we will consider. The first type of error was the simple failure to recognise a familiar person: 114 such incidents were recorded. The explanation for such errors is presumably the failure of the perceived face to access the internal stored representation of the familiar person's appearance. Someone you know but fail to recognise may, for example, have cut their hair and shaved off a beard, or lost weight dramatically, or may have aged 15 years since you last saw them.

The second type of error reported was the misidentification of one person as another (314 incidents). Such errors tended to be short-lived and to occur under poor viewing conditions (e.g. a brief glimpse of someone). Usually, the misidentification took the form of mistaking an unfamiliar person for a familiar one (thinking the person walking towards you is your granny then realising she is a stranger). The similarity is sufficient to momentarily activate the stored representation of the familiar face, though a second and better look reveals the discrepancies.

Young et al. (1985) collected 233 reports of a third type of error. This involved seeing a person, knowing she is familiar (i.e. that you have seen her somewhere before), but being quite unable to think who she is, where you know her from, or what her name is. Typically this happened with slight acquaintances (rather than close friends or relatives) encountered out of their usual context—for example, seeing a clerk from your bank out shopping in the street.

The fourth and final type of error is the inability to remember someone's name (190 incidents). In over 90% of these instances, the diarists reported being fully aware of who the person was in the sense of what her occupation was (99%) and where she was usually seen (92%), but the name remained elusive. There seems to be something different, and something difficult, about names.

Young et al. (1985) interpret the fact that you may be able to recall all you know about someone except their name as suggesting that names may somehow be stored apart from the other information you possess about familiar people. They argue that satisfactory face recognition requires the involvement of at least three separate mental systems. The first is the *face recognition system,* in which the stored representations of familiar faces are held. The second is a *semantic system,* in which is located all the general knowledge you possess about people you know. Third and last is a system from which the spoken forms of words, including names, are retrieved.

Drawing upon the findings of Young et al. (1985), and much other research on face recognition, Bruce and Young (1986) proposed the model of face recognition illustrated in Fig. 1.18. We will concentrate initially upon the right-hand side of the model. Each box represents a separate processing module or store and the arrows indicate transmission of information between these modules.

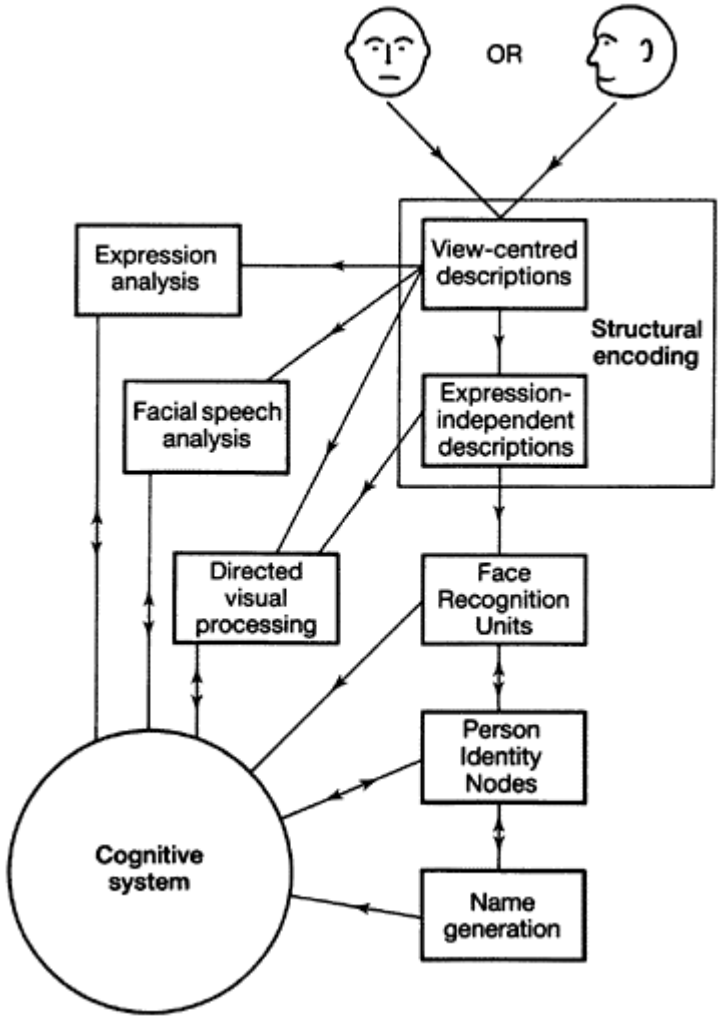

FIG. 1.18 Bruce and Young's (1986) model of face recognition (redrawn from the original).

The model begins by postulating that structural descriptions of the face are derived, first view centred, as in Marr's $2\frac{1}{2}$-D sketch, then as 3-D representations that are *independent* of facial expression. The independence is introduced partly because we recognise faces of familiar people independently of whether they are smiling or frowning, but also because there is neuropsychological evidence, reviewed by Bruce and Young (1986), that expressions are processed separately. So, some patients with neurological damage can correctly identify faces but not their emotional expression, while other patients show the

reverse effect. Therefore, Bruce and Young included a separate *expression analysis* module in their model.

The structural encoding stimulates the *face recognition units*. Each of these units contains stored details of the face of a known person. The closer the newly seen face is to the stored details in any face recognition unit, the higher the activation in that unit will be. The face recognition units are linked both to the *cognitive system* and to *person identity nodes*. The cognitive system is the store of information about the individual and is therefore a part of semantic memory (see Chapter 9). So, when the face recognition unit for a familiar face is activated, for example by seeing Marilyn Monroe, that makes available from the cognitive system semantic information about Ms Monroe, such as the fact that she was a famous film star. The *person identity nodes* are the point at which person recognition is achieved. They receive input not only from face recognition units but also from an analysis of voices, names, posture, clothing, etc. Such a level is included in the model because recognising a person does not require seeing his or her face. It might come about through hearing his or her name, or his or her voice, etc. Only when the person identity node has been sufficiently stimulated by input from one or more of its sources is the person recognised. Only after such recognition can the name of the person be generated, hence the final *name generation* box.

The structure of this model captures the data reported by Young et al. (1985). Where errors were found in that study, it is possible to locate them in the failure of processing within the Bruce and Young model. So, for example, recognising that a face is familiar but not knowing any more about the person implies that a face recognition unit has been activated but that, for some reason, this has not stimulated the cognitive system or the person identity node sufficiently to retrieve more information about the person. A failure to recall a name for someone about whom one can recall many other details implies problems at the final stage between the person identity node and the name generation system.

Two remaining modules require a brief explanation. Bruce and Young included a *directed visual processing* module because they point out that we can, strategically, actively direct our attention to process certain aspects of a face. We may look for particular features; for example, if we know that the friend that we are to meet at the station has long blonde hair we may look specifically for such hair, among the distant alighting passengers. The *facial speech analysis* module is included because lip-reading has been shown to be a separate cognitive ability, independent of face recognition (Campbell, Landis, & Regard, 1986). Some individuals who have suffered brain damage can recognise faces but not lip-read, while others may lip-read but be unable to identify familiar faces. So facial speech analysis seems to be a separate system.

AN INTERACTIVE ACTIVATION MODEL OF FACE RECOGNITION

While models such as those of Bruce and Young (1986) are important in clarifying theoretical ideas, even greater power comes from models that can be formally simulated by computer programs. Such simulations usually highlight any lack of specificity in the modelling and can reveal properties

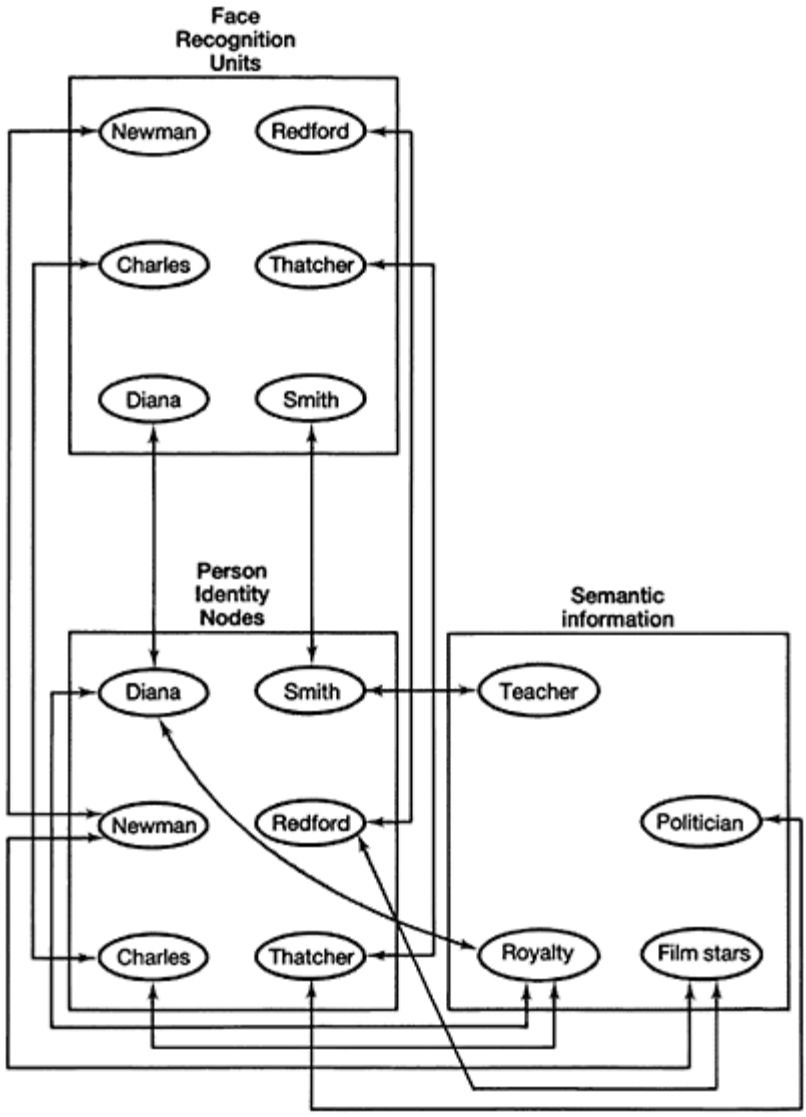

FIG. 1.19 The interactive activation model of face recognition (redrawn from Burton et al., 1990).

of the model that were not obvious even to the authors of the model. Burton, Bruce and Johnston (1990), therefore, developed a computational model based upon the Bruce and Young (1986) framework.

Burton et al.'s (1990) model has three separate "pools" of units, as has Bruce and Young's model. These contain, respectively, units for face recognition (FRUs) for each

individual's face, person identity nodes (PINs) for each person "known" to the model, and semantic information known about the individuals. These units are linked in appropriate ways (see Fig. 1.19). There are FRUs and PINs for each person represented in the model, and these are connected to appropriate semantic information. So, for example, there would be a FRU for Ms Monroe's face, a PIN for her and links from the PIN to semantic information about her, such as her being dead, having been a film star, etc. Unlike Bruce and Young's model, semantic information is linked to PINs but not to FRUs. One result of the formalising of the model was to demonstrate that the FRU → semantics link was inappropriate.

In the model, recognition occurs if the activation in the relevant PIN reaches a given threshold. Where units are linked they will transmit excitation to other units to which they are linked if they themselves are excited. As with all such models (see Chapter 2), there are inhibitory links between units to counterbalance this excitation. So, whenever one PIN increases in excitation, this is transmitted through *inhibitory* links with all other PINs and will *reduce* their activation.

Burton and et al.'s model has been very successful in simulating a wide range of experimental findings about face recognition. We will take two examples, first explaining the research findings that must be simulated by an adequate model. One example is semantic priming. When the face to be identified is preceded by the face of someone with whom the person is associated (e.g. a picture of Stan Laurel preceded by one of Oliver Hardy), the recognition of the face as familiar is quicker than if the preceding face is of an unrelated but familiar person (e.g. Bruce & Valentine, 1986). Burton and coworkers' model simulates this priming by producing easier recognition when an associated face follows seeing a given face than when the second face is unconnected with the first.

Burton et al.'s model also simulates the *distinctiveness effect*. It is well established that distinctive familiar faces are more quickly recognised than more typical faces (e.g. Valentine & Ferrara, 1991). It is also frequently reported that recognition of faces from one's own race is easier than for other races. Why should this be so? Valentine and his associates (e.g. Valentine, 1991; Valentine & Endo, 1992) have developed a theoretical framework which proposes that faces are encoded as points in a multi-dimensional space, where the dimensions represent those upon which the faces are encoded. Decisions on whether a face is familiar, in Valentine's model, are based upon analysing the new face on the multiple dimensions, determining the distance of this new face from its nearest neighbour. Distinctive faces are more dissimilar to other faces across these multiple dimensions. Valentine and Endo (1992) are able to explain the finding that it is easier to recognise faces from one's own race than those of other races by assuming that the dimensions upon which faces are encoded are refined over a lifetime's experience, but that the important dimensions for recognising, say, an Afro-Caribbean face, may not be those which best discriminate between Chinese faces.

These distinctiveness effects are simulated by Burton and et al.'s model. It assumes that distinctive faces share less features than do typical faces. Typical and distinctive faces were modelled with typical FRUs receiving inputs from features shared with several other faces, while distinctive FRUs had features shared with few other faces. When the model was run, the PINs for distinctive FRUs were more quickly and strongly activated than those for typical faces.

We have illustrated how the model of Burton et al. is able to simulate two important features of face recognition. One unexpected outcome from developing the simulation was that an explanation for the poor recall of names emerged that did not require the assumption of a separate name store, as in Bruce and Young's model, somewhere at the very end of the processing. The "name effect" is a robust finding, and it occurs not just within the diary data of Young et al. (1985). For example, McWeeny, Young, Hay and Ellis (1987) showed that people take longer to put names to faces than occupations to faces, even when the names are actually occupations and the *same* words are used as occupations or names, for example "Mr Baker" or "is a baker".

Burton and Bruce (1992) pointed out that, unlike most semantic properties associated with an individual, a person's name is connected to only one PIN. So, while the semantic information that X is a doctor, a parent, lives in Blackpool and so on is likely to be shared with other known individuals, the fact that her name is Marigold Beckett is likely to be a unique link between the name and the person node. When this was represented in simulation runs of Burton et al.'s model, it was always found that names received the least activation and were slowest in reaching their maximum activation of *all* the semantic information units. All the semantic information was activated before the name was retrieved. Thus, the simple interactive simulation model was able to show that what had for so long seemed a puzzling feature of recognising a person, and had led to specific boxes being added to the model, was really a natural consequence of the nature of the associative network linking the information about any individual. Names are special only in so far as they are special and distinctive for an individual. It is not necessary to assume that they are treated in a different manner to other information about an individual within the cognitive system.

THE ROLE OF CONTEXT IN OBJECT RECOGNITION

If you met your grandmother outside her house, you would almost certainly immediately recognise her. However, if you passed her in the street in a different town that you did not expect her to be visiting, you might miss her, or doubt that the person really was her. Object recognition is clearly determined in large measure by the visual description of an object, but objects are also normally perceived in contexts. Does the context in which an object appears affect the speed or accuracy with which it can be recognised for what it is?

The subjects in Palmer's (1975) experiment were asked to identify briefly presented line drawings of objects. On some trials, they were first given a picture of a scene (e.g. of a kitchen scene) to inspect. The briefly presented drawing which followed could then be either appropriate to the preceding scene (e.g. a loaf of bread) or inappropriate (e.g. a drum). Accuracy of object identification was highest when the object picture followed an appropriate scene, and lowest when it followed an inappropriate one. With no preceding scene, accuracy was intermediate. We may say, then, that the context provided by an appropriate scene *facilitates* the subsequent recognition of a briefly seen object, whereas that provided by an inappropriate scene *inhibits* subsequent recognition.

As described in the previous section, experimenters have asked whether one person's face is more easily recognised if it is seen immediately after the face of a second person with whom the first is associated than if it is seen after the face of an unassociated person.

"Association" here means pairs of people who share a common occupation and/or tend to be seen together. Bruce and Valentine (1986) showed that where two famous faces occurred together in the sequence, the time taken to respond to the second face was less if the first face was associated than if the first face was unrelated.

It would appear, then, that the recognition of an object can be facilitated by the context in which it is seen—either the general background context as in Palmer's (1975) object recognition experiment, or the other objects recently recognised as in Bruce and Valentine's (1986) face-priming experiment. But *how* can context influence recognition? We have already argued that recognition occurs when an analysed pattern accesses and activates a stored representation. One possibility proposed by Seymour (1973) and Warren and Morton (1982) for objects, and by Hay and Young (1982) and Bruce (1983) for faces, is that context works by contributing some activation to the representations for patterns that the context suggests are likely to be perceived. Representations that are already partially activated from within by the context will then require less input from the stimulus pattern to be fully activated and to trigger recognition. Inspecting a kitchen scene will, on this account, cause partial activation of the stored representations of the appearance of objects likely to be encountered in a kitchen (loaves of bread, forks, casseroles, cookers, etc.). Each of these visual patterns will be recognised more easily as a result of this priming than if it suddenly intruded into, say, a jungle scene. Similarly, if Nancy Reagan's face appears on the television, you will recognise her more easily if you have recently recognised Oliver Hardy.

SUMMARY

Recognising a familiar face must involve comparing a perceived stimulus pattern against a set of stored representations. Simple template and feature models have problems coping with the variability of natural patterns (e.g. familiar faces in different orientations, with changing hairstyles, etc.). Where faces are concerned, the visual system appears to try to counteract the variability of external features like hairstyle by an increasing reliance on internal features as the basis of the recognition of familiar faces.

Faces and objects are normally encountered against a complex, changing visual background. Psychologists have made some progress in identifying the cues that enable elements of a visual array to be grouped together as parts of the same object. Such grouping is a necessary preliminary to recognition. In particular, the work of Marr (1982) not only suggests possible stages through which object recognition may be accomplished, but also clarifies the types of explanation that need to be sought when attempting to understand any part of the cognitive system. The cues underlying distance and movement perception have also been analysed in some detail.

It is commonly assumed that word, face and object recognition converge on a common "semantic" stage at which is held knowledge of the meanings and uses of words, the uses and properties of objects, and the characteristics and personalities of people. A model was presented to account for the interconnection between the cognitive processors responsible for different types of recognition, for comprehension and for naming, using as sources of relevant evidence the recognition problems experienced by both normal and brain-injured individuals.

Fuller accounts of visual perception will be found in V. Bruce and P. Green (1990), *Visual perception: Physiology, psychology and ecology,* 2nd edn, and in G.W.Humphreys and V.Bruce (1989), *Visual cognition: Computational, experimental and neuropsychological perspectives*.

2
Reading Words: Sight and Sound in Recognising Patterns

Imagine that one evening you arrive home late with a friend and on the kitchen table you spot a note addressed to you. You pick the note up and read: *Please wash the potatoes and put the casserole in the oven.* Now you might object to this request, but you do not find it difficult to decide what "potatoes" means. A 7-year-old child, on the other hand, might have considerable difficulty in deciding what to wash and what to put in the oven. Anyway, having read the note, you both decide to have a drink and put your feet up for a while before starting to prepare the meal. This is a simple scenario and ones like it must happen thousands of times each day, yet the processes that allow us to read the note, to remember its content, to develop intentions about future actions, to remember those intentions and so on are taken for granted. Much of this book is aimed at showing that although these actions seem easy to perform and to require little conscious effort, that does not make them easy to explain. Likewise, at a number of points we will see how simply reflecting on our subjective experiences of reading is not always a reliable guide to how we actually do it.

Reading has been extensively studied by psychologists, partly because it is a skill that requires some effort to acquire (even if the effort is then soon forgotten), and partly because it is a specifically human, culturally transmitted cognitive activity. When Huey (1908) reviewed the first 10–15 years of experimental research on reading, he remarked that: "to completely analyse what we do when we read would almost be the acme of a psychologist's achievement, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilisation has learned in all its history". Psychologists have not yet finished unravelling the complexities of that remarkable performance, but they have made a start.

Where should the study of reading begin? One obvious place might be the recognition of letter shapes, since letters are the building blocks of written words. But there is an even earlier stage in reading. Before you can begin to recognise a letter or the word in which it belongs, your eyes must first alight on the word. The eye movements made during reading are quite subtle and we begin with a discussion of them. Later in the chapter, we consider the processes by which a written word evokes its meaning and sound in the reader's mind.

SCANNING THE NOTE: EYE MOVEMENTS AND TAKING IN INFORMATION

As you read the note about the potatoes and the casserole, you were probably unaware of your eye movements. Even if you had been attending to them—perhaps having read a chapter like this earlier in the day—you may not have had an accurate feeling for what was going on. However, had your companion observed you while you were reading the note, he or she might have noticed that your eyes did not move smoothly along the line of print but progressed by a succession of quick flicks interspersed by moments when your eyes were quite still.

The stationary moments are called *fixations* and the flicks are referred to as *saccades* (the term used by the French ophthalmologist Javal who first reported them in 1887). Saccades and fixations are not unique to reading; they happen whenever we inspect a static scene. In fact, the only time we can move our eyes smoothly is when we are tracking a moving target. You can easily verify this by asking a friend first to fixate the tip of a pencil while you move it in a straight line, and then to move his or her eyes in a straight line without a target to follow. In the first condition, you will observe smooth, continuous eye movements; in the second, you will see saccades and fixations.

Figure 2.1 shows a passage of text and superimposed upon it a somewhat idealised representation of the eye movements that a skilled reader would be expected to make. The circles above the words represent fixations, with the diameter of the circle indicating their likely durations (larger circles for longer fixations). The arrows show the saccades. Note that most saccades are forwards, but in reality some 10–20% of eye movements (excluding those from the end of one line to the beginning of the next) go backwards to re-fixate words which have been fixated once already. These are called *regressions*.

The average fixation of a skilled reader lasts 200–250 msec (that is, between one-fifth and one-quarter of a second). The average saccade takes only 25–30 msec (a mere one-fortieth of a second). Compare these two figures and you will realise that a reader's eyes are in fact still—that is, fixating—for about 90% of the time during reading. It may *feel* as if your eyes are constantly moving when you read, but they are stationary for most of the time, which illustrates how misleading one's subjective impressions can be.

All of the uptake of useful information in reading occurs during fixations. We have known since the work of Dodge (1900) and Holt (1903) that little or nothing is perceived during a saccade (a phenomenon known as "saccadic suppression"). Even a strong flash of light is unlikely to be perceived if it occurs entirely within a saccade (Latour, 1962). Suppression also extends into the first 60–80 msec of a fixation: both in reading and viewing a scene, it takes that long after a saccade before information begins to be utilised (McConkie, 1983).

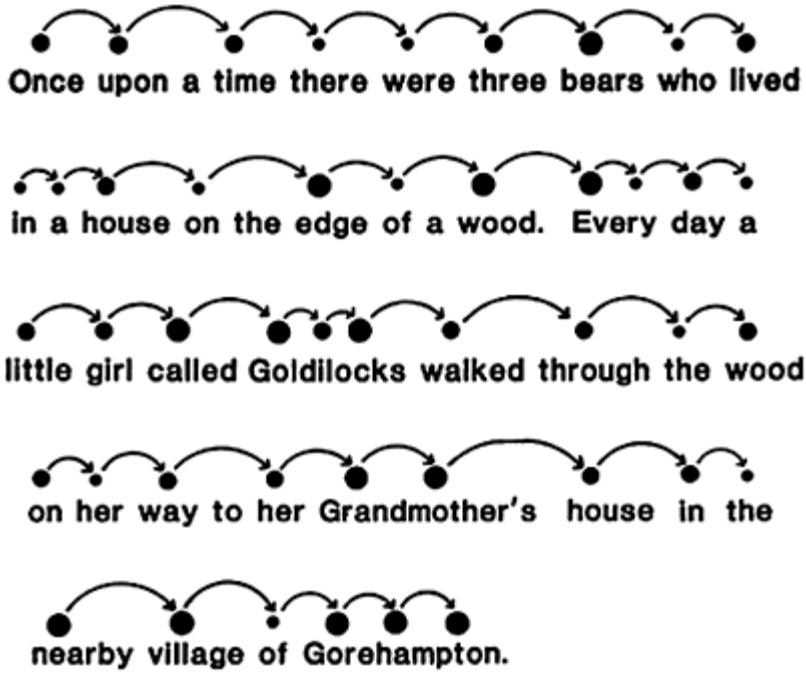

FIG. 2.1 The pattern of eye movements that a skilled reader might make in reading a passage of text. Circles represent fixations, with larger circles indicating longer fixations, and arrows represent saccades. (For simplification, we have omitted the 10% or so of backward, "regressive" eye movements that occur in natural reading.)

But why is it necessary to make these eye movements at all when reading? Why is it not possible to read a whole page of a book or a whole notice by giving it one long fixation in the middle? One of the main reasons is that the quality of visual information picked up declines rapidly with distance from the point of fixation. Detailed vision is made possible by light-sensitive cells called *cones,* which are located in the retina at the back of the eye. The cones are most densely packed at the central region of the retina, the *fovea,* and this allows greatest visual acuity. When one fixates on a point, light from that point impinges on the fovea. Moving away from the fovea, the density of cones decreases, bringing with it a corresponding decrease in visual acuity away from the fixation point. Reading requires reasonably high visual acuity. If letters or words are

briefly displayed at various positions in the visual field, then the likelihood of being able to identify them correctly declines rapidly with distance away from the fixation point.

Although acuity declines more or less symmetrically around the fixation point, information uptake in reading is not symmetrical. One of the most important findings showing this lack of symmetry refers to something called the "perceptual span". The perceptual span is the area of text around the fixation point from which useful information is picked up. When reading normal English text, this area extends further to the right than it does to the left. Rayner, Well and Pollatsek (1980) suggest that the span is up to 15 characters to the right but only 3–4 characters to the left (see Fig. 2.2). One obvious reason for this asymmetry in English might be that information to the left of fixation has already been processed (equally obvious is that information to the right of fixation has not been processed—both factors probably contribute to the asymmetry). Unlike visual acuity, the asymmetry of the perceptual span is constrained by our learning experiences rather than by anatomy. Pollatsek, Bolozky, Well and Rayner (1981) showed that for readers of Hebrew, which runs from right to left, the perceptual span is biased to the left of the fixation point. Inhoff, Pollatsek, Posner and Rayner (1989) asked readers of English to read English texts that were printed from right to left rather than the usual left to right. After a short amount of practice on these reversed texts, Inhoff et al. showed that the asymmetry of the readers' perceptual span reversed, so that more letters were picked up to the left of the fixation point than to the right. This suggests that the form of the perceptual span is guided by our reading experiences but remains very flexible.

Readers tend to fixate on a point somewhere between the beginning and middle of a word (Rayner, 1979). This point is referred to as the preferred viewing location. However, this point is not always the optimal viewing location, which is the point in the word from which viewers can obtain most information about the word. In English, this optimal point is usually slightly to the left of the centre of the word. O'Regan and Jacobs (1992) showed that deviation from the optimal viewing position, as measured in number of letters away from the optimal point, slows down reading of a single word to the extent of 20 msec per letter.

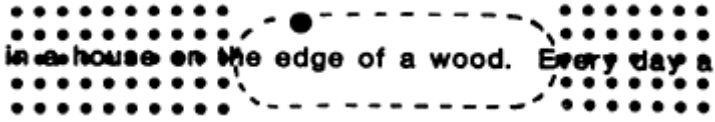

FIG. 2.2 The "perceptual span" in reading typically extends further to the right of the fixation pont than to the left.

The discrepancy between the preferred and optimal viewing positions is attributed to two factors: inaccuracy in the eye movements themselves and something called preview benefit (Rayner & Morris, 1992). The region of text which is in view but which does not fall on the fovea falls on an area known as the parafovea. For example, in Fig. 2.2, if you fixate on the **d** in **edge** at a normal reading distance, then the word **wood** falls in the

parafovea. Experiments have shown that if a word has been "previewed" in the parafovea, then when it is fixated upon the duration of that fixation is shorter than if the word had not been "previewed". The argument is that the preview gives some information in advance and so the point to which the eyes then move need not correspond to the optimal point. Exactly what information is gleaned from preview is a matter of some debate, but whatever it is it clearly has an influence on the pattern of fixations.

Fixating on a word is only the first step in reading it. If you fixate a word in a foreign language, particularly one in an unfamiliar alphabet like Russian or Arabic, then you are unlikely to make much sense of it. If the word is from your own language, however, the word will activate its meaning and its sound without apparent effort. How is this achieved? What happens inside your head that distinguishes between fixating a word that is written in your own language and fixating on a word that is part of a language that you cannot read? If we could answer that question completely, we would have reached Huey's "acme". We haven't yet succeeded in doing so, but we can at least sketch the outlines of a plausible answer.

A PRELIMINARY ACCOUNT OF THE RECOGNITION OF FAMILIAR WRITTEN WORDS

Recognising Letters

Many theories of visual word recognition assume that at least some of the component letters of a written word must be identified before the word itself can be recognised. By "identify a letter" we do not mean name it. Here, identification refers to the visual categorisation of a stimulus pattern such as a letter as an exemplar of a known type, in this case as a particular letter of the alphabet. Coltheart (1981) reported the case of a man who, following brain injury, could understand most written words, but could pronounce very few, and was rarely able to read aloud invented nonwords like ANER. His letter naming was also poor, yet he could reliably respond "same" to pairs of nonwords like ANER/aner and "different" to pairs like ANER/aneg. These decisions were possible because the man could still identify and compare the component letters of words or nonwords visually, though he could no longer pronounce or name them.

Letter recognition is commonly assumed to be a necessary part of word recognition, but we have as yet only an ill-defined notion of how exactly it might be achieved. An early theory proposed that we might possess internal letter *templates* against which stimuli could be compared. You can think of these templates as stencils, which you could place over a letter to check for a match. The problem with this theory is evident from the limitations of machine systems that employ this mode of letter recognition. An example is the computer reading of the lettering on cheques. Those rather odd letter shapes are designed to be minimally confusable (at least to the computer), and the systems that recognise them are very intolerant of even slight changes in letter shape. Yet as Fig. 2.3 illustrates, even within upper or lower case, the same letter can take a variety of different, if related, forms. Recently, Loomis (1990) has produced a complex model of character recognition, implemented on a computer, that makes use of templates. In a series of experiments on visual and tactile stimuli, he compared the performance of his model to

that of people. The task of the subjects in his experiments was to learn sets of characters, many of which were entirely novel. Not surprisingly, people found it harder to learn the characters in some sets than in others. When tested on the same series of character sets, Loomis' computer model had most difficulty distinguishing characters within those sets which people also found most difficult, while having least difficulty with those that people also found least difficult. Loomis is careful to point out that this does not mean that his model would also mimic people's performance on letter recognition tasks involving other measures of performance, such as reaction times, but the model's success on discrimination tasks at least suggests that there could be some role for templates in letter recognition. However, it is fair to say that the model came nowhere near dealing with the type of variation of the same character that is found in the vast array of different scripts and hand-writings that we encounter as readers.

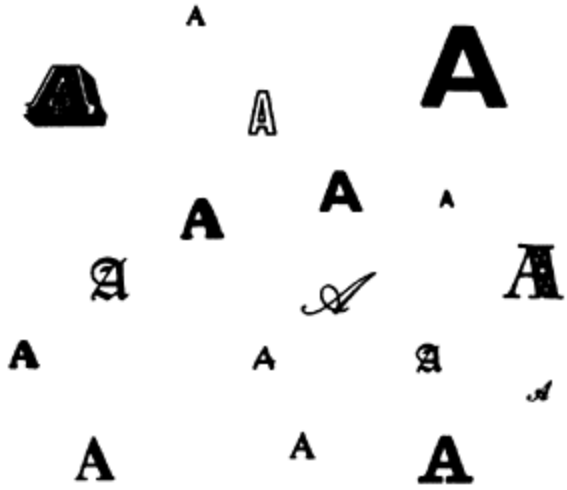

FIG. 2.3 Any template-based model of letter recognition would have great difficulty coping with variations in letter shape and size.

As an alternative to templates, some theorists propose that letters are recognised in terms of sets of *distinctive features* (Oden, 1979). A distinctive feature is a property or aspect of an object that helps to distinguish it from other objects. The distinctive features of letters might include lines and curves in varying orientations. Thus, the features for the letter A might be [right-sloping oblique], [left-sloping oblique], [horizontal line]. That would not be enough, however, since there are many ways you could combine two sloping and one horizontal line without forming a letter A (see Fig. 2.4). The "features" therefore would need to include specifications of the *configuration* that the elements must adopt in order to be an acceptable letter. Some theorists have suggested that there are separate stages for locating and integrating features of objects such as letters: the first stage happens in parallel and the parts or features are located in the visual array, and it is