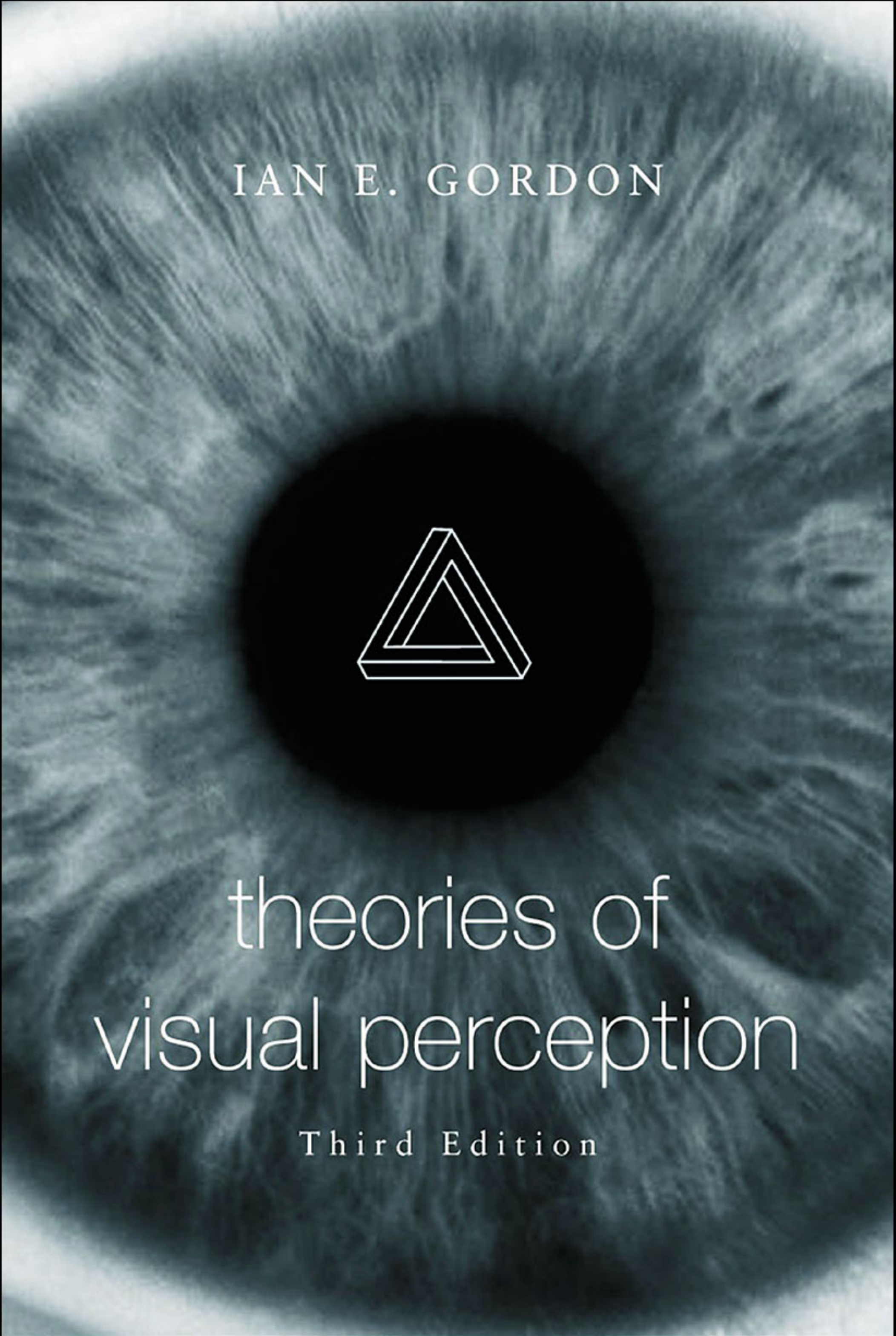


IAN E. GORDON

theories of visual perception

Third Edition

Theories of Visual Perception

This thoroughly revised and expanded edition of *Theories of Visual Perception* contains critical accounts of the major approaches to the challenge of explaining how we see the world. It explains why approaches to theories of visual perception differ so widely and places each theory into its historical and philosophical context.

Theories of Visual Perception begins with a discussion of the hierarchical status of theories of perception and the impact of technological and methodological developments, giving the reader a clear understanding of the background to the subject. It provides accounts of the major theories of visual perception, ranging from early theories by some of the most influential writers in perception, such as Helmholtz and the Gestalt School, to recent work involving computer models of vision and cognitive brain imaging. Chapters cover subjects including Brunswik's probabilistic functionalism, Gibson's theory of direct perception, neurophysiology and empiricism, providing an in-depth analysis and critical appraisal of each theory.

This volume provides extensive coverage of the most significant theories of visual perception, along with a valuable insight into their historical backgrounds. It will be welcomed by students and researchers in the fields of psychology, physiology and neuroscience.

Ian E. Gordon has taught and carried out research in universities in Australia, Canada, New Zealand and Tianjin (where he is visiting Professor). He has published numerous articles on perception and has also written on psychology and art.

Theories of Visual Perception

Third edition

Ian E. Gordon

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK

First published 2004
by Psychology Press
27 Church Road, Hove, East Sussex BN3 2FA

Simultaneously published in the USA and Canada
by Psychology Press
270 Madison Avenue, New York NY 10016

This edition published in the Taylor & Francis e-Library, 2005.

Psychology Press is part of the Taylor & Francis Group

This publication has been produced with paper manufactured to strict environmental standards and with pulp derived from sustainable forests.

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

Library of Congress Cataloging in Publication Data
Gordon, Ian E.
Theories of visual perception / Ian E. Gordon.
p. cm.
Includes bibliographical references and index.
ISBN 1-84169-383-9
1. Visual perception. I. Title.

BF241.G67 2004
152.14–dc22 2004006844

ISBN 0-203-50225-6 Master e-book ISBN

ISBN 0-203-59534-3 (Adobe eReader Format)
ISBN 1-84169-383-9 (Print Edition)

For Ross Day, scholar and enthusiast

Contents

Preface to this third edition

Responses to the first two editions of *Theories of Visual Perception* were generally very positive. Students in five countries told the author that the book helped them to grasp the essentials of some of the various theories of visual perception. Many lecturers have adopted the book. With two exceptions, the book received favourable reviews. A number of young researchers in artificial intelligence have reported that *Theories of Visual Perception* was their entry point into vision. All this has been very gratifying.

Any feelings of complacency were quickly checked by some general criticisms from the author's own students at Exeter University. In the main, these focused on two chapters: that on psychophysics, and that on Brunswik's theory. To argue that the essential nature of a sensory threshold has been a subject of theoretical debate for 100 years, and that a major branch of visual research had arisen because of this debate, did not convince these young readers. After much thought, the original chapter on psychophysics has been dropped from this volume.

Brunswik's probabilistic functionalism struck many students as a real oddity. They understood the chapter on his theory, but found his own writings difficult in the extreme. The chapter has been retained in this second edition for two reasons: first, because of the present author's hunch that Brunswik may have been right in his intuitions concerning the basic nature of perception; second, because of some recent publications concerning perception and the statistical nature of real world events. This new work has been described in the present chapter on Brunswik's contribution.

Since the second edition of this book appeared, there has been a torrent of recently published research findings concerning visual perception. At this point it must be stressed that *Theories of Visual Perception* is not a textbook on visual perception *per se*. Rather, the book is an attempt to describe how a number of general theories of visual perception developed: their backgrounds, their underlying assumptions, their strengths and weaknesses, and their current status. For those wishing to read fuller accounts of experimental findings in vision research, there exist a number of remarkably good textbooks, one example being Bruce, Green and Georgeson (1996), which is the present author's favourite. Readers should also enter the Web for up-to-date findings in particular areas – search engines, such as Google, can be invaluable.

This is the place to say something about the style of *Theories of Visual Perception*. When acting as a book reviewer, the present author used to take a quick glance at the References section in his review copy. This gave him a feel for the likely coverage in the book and how up-to-date this was. Anyone reading the present volume who repeats the above tactic will be struck by the large number of references to work published around the middle of the last century – a lifetime ago. There are two reasons why this is the case. The first has to do with the present author's teaching style. The second is historical.

Different academics have different approaches to the challenge of teaching, and there seems to be no single best one. All agree, however, that good teaching should not comprise the stuffing of students' heads with details. What is important is to give students a framework within which they can organize their studies. As a lecturer, the present author believed that by showing students the origins of a particular theory – its philosophical and historical background, its initial emergence and development, together with the first criticisms of the theory – students would gain an understanding that would enable them to grasp with ease subsequent developments or refinements of the theory.

The second reason for including much material published in the middle of the twentieth century is simply this: as we shall see, this is the period during which many theories of perception began to take shape. At the same time, psychologists began to be much more self-conscious about theories *per se*. Terms such as 'hypothetical construct' and 'intervening variable' were appearing in the literature for the first time. Theorizing was being taken seriously and on a scale that included memory, learning, language and perception.

The organization of this third edition will now be described.

Chapter 1 is an introductory chapter and begins with a general discussion of the hierarchical status of scientific theories in general and theories of perception in particular. The author has become increasingly aware of the importance of technological/methodological developments in scientific progress, particularly the impact on empirical research and subsequent theorizing. Examples of this impact are given.

Chapter 2 is an expanded account of Gestalt theory. The chapter now contains much more material on perception by newborn infants. Some of the startling discoveries of recent years are increasingly supporting the Gestalt psychologists' nativist stance concerning the origins of perception. A review of this research on infants reinforces the claims made in Chapter 1, concerning the contribution of improved techniques in discovering new facts about perception. The chapter also contains an extended account of algorithmic information theory. The reason for this is that this new branch of mathematics offers a new way of measuring simplicity. The central claim of

Gestalt theory was that perception strives towards simplicity. There is now good reason to hope that this claim can now be tested.

In *Chapter 3* on Brunswik's probabilistic functionalism, an account of a new theory – the empirical theory of vision – has been added and the remarkable new techniques used to explore the implications of this theory are described in some detail. The chapter concludes that Brunswik's original claim concerning the statistical nature of perception may have been correct.

Chapter 4 on neurophysiology and perceptual theories now contains descriptions of the latest brain-scanning techniques, by which cortical functions can be studied in conscious volunteers as they perform a variety of tasks.

Chapter 5 on empiricism is largely unchanged, apart from a discussion as to what will happen if supporters of this constructivist approach are forced to accept the serious implications arising from recent discoveries in infant research. These are showing that the newborn infant possesses a number of impressive perceptual skills. A possible future role for empiricism is discussed.

Chapter 6 on Gibson's theory of direct perception is also largely unchanged, apart from the inclusion of some examples of the most recent findings in this area. Generally, this was the best-liked chapter in the previous editions of *Theories of Visual Perception* and the author lacks the ability to improve on it. However, it seems fair to say that, in very recent years, only one potential theoretical advance has taken place within this tradition, although empirical research continues apace. A possible resolution of the conflict between direct perception and constructivist theories (the theoretical advance referred to) in terms of different neural pathways has been included. The other main additions to this chapter are accounts of the adoption of Gibson's theories by some of those working on problems of design and also robot vision; these are unusual and unexpected developments.

Chapter 7 on computational theory is also largely unchanged, except that it is now more critical of Marr's approach and includes a description of how some recent theorists appear to be moving away from Marr's position.

Chapter 8 offers some final remarks on theories of visual perception.

Acknowledgements

Several people have offered invaluable help during the preparation of this third edition. Thanks are due to my agent Michael Coombs for his encouragement and hard work on my behalf. My colleagues at Exeter University, Dave Earle and Alan Slater, were, as always, generous with their time: their expertise has been an important factor in deciding what to include in this new edition. Dr Peter van der Helm of the University of Nijmegen stimulated my interest in algorithmic information theory and sent me a pre-print of his important 2002 paper.

Thanks are due to Tessa and Harry Gordon for all their help and to my friend Chrissie Morris for commenting on many parts of the manuscript and for her constant encouragement.

Bettina Newman prepared the illustrations with her usual skill and efficiency.

Ian Gordon
Exeter University, February 2004

1 Theory and method

In this chapter we will offer some general remarks on the nature of scientific theories – a reasonable beginning to a book bearing the present title. Then we shall say a little about the role of new techniques in the discovery process and subsequent theorizing.

The calibre of scientific theories

Much has been written about the nature of scientific theories. Perhaps the best-known writer in this field is Karl Popper (1902–1994). Popper argued that no scientific theory can be proved to be correct; it can only be shown to be wrong, or at least flawed. It is the job of scientists to find ways of challenging theories by empirical means. This is only possible if theories are designed to be open to empirical validation – they must be falsifiable. That the moon is made of green cheese is a silly theory (or hypothesis), but it is at least falsifiable: we can go there and check. In contrast, Freud's psychoanalytic theory is not silly, but many have doubted whether, for example, his concept of the Oedipus complex could ever be falsified by empirical observation.

Popper (1959) demonstrates how the process of empirical testing leads to an evolution of theories whereby earlier theories become corrected and often embedded within newer, more inclusive ones. Thus, Newton's theory can still be used in many contemporary situations, such as sending a rocket to the moon. But it was the presence of certain weaknesses in the theory (facts it could not explain) that led to Einstein's theory of relativity – which includes Newton's laws as special cases.

It is now generally agreed that when Popper made his assertions he was thinking about what can be called 'great theories'. Examples are Newton's laws, Darwin's theory of evolution, relativity theory and, most recently, the quantum electrodynamic theory of light. There is no need to write at length about these theories. However, here are a few examples of their power.

Newton's formula, $G = (m_1 \times m_2)/d^2$, where G = gravitational force, m_1 and m_2 are the masses of two bodies and d^2 is the square of the distance between them, has been described by the distinguished physicist Richard Feynman as the most powerful equation of all time (Feynman, 1999): it can be used to predict the movements of the planets around the sun, the orbit of the moon around the earth and even the heights of tides – and these are but a few examples of its power.

Darwin's theory of evolution has been described by the philosopher Daniel Dennet as, 'The best idea anyone ever had' (Dennet, 1991). The theory explains the evolution of all species and has survived every test to which it has been subjected. For example, in the 140 years since the theory's publication no single fossil has ever been found in the 'wrong' geological stratum.[1]

Einstein's famous equation, $e = mc^2$, where e = energy, m = mass and c = the speed of light, led eventually to the splitting of the atom: a landmark in the progress of knowledge.

The quantum electrodynamic theory is hard for the non-physicist to grasp. However, one fact demonstrates the theory's power: the theory has been subjected to a series of ever more stringent experimental tests. For example, a number of experiments have found the value of what is known as 'Dirac's number' to be 1.00115965246 with an uncertainty of around 4 for the last digit. The quantum electrodynamic theory predicts the value to be 1.00115965246 with a larger uncertainty for the last digit. If the distance from New York and Los Angeles (about 3000 miles) were to be measured to this degree of accuracy, the above mismatch would amount to the thickness of a human hair. This is a *very* powerful theory.

What we may call 'good' theories fall short of the extraordinarily high standards exhibited in those outlined above. They may explain some important phenomena, but they give rise to predictions that may not always be confirmed by empirical testing. Examples include: Mendelian genetics, Marx's theory of the historical process, Keynes's economic theory, Chomsky's theory of syntax and, closer to the theme of this book, the Young–Helmholtz theory of colour vision.

There are no great theories in this book. It will be claimed in later chapters that Marr's work and some physiological discoveries have led to 'good' theories, in the sense outlined above. However, most of the work that is to be described might better be described as coherent sets of ideas – ideas that have in fact prompted much high-quality research. We may call these 'utilitarian' or 'working' theories. For example, there are many perceptual researchers who believe that the essence of visual perception is that it is a knowledge-driven process (or sets of processes); in other words, perception is essentially a constructive process. Other workers have assumed that perception is largely the product of innate brain processes. Currently, there are many who claim that perceptual processes cannot be adequately understood until the intimate relationship between perceivers and the environment(s) in which they evolved has been made the focus of major research programmes.

---

[1] By a nice coincidence, a new species has been seen to emerge in the month when this is being written. It is a plant growing on the river bank in the city of York.

These sets of ideas have not much in the way of formal structure. Few of them are capable of generating quantitative predictions and, as we shall see, even when flaws are found and awkward facts are discovered, this does not lead to the abandonment of the approach or even to major alterations to the theories. What such theories have done is, first, to provoke ingenious experiments, second to unite like-minded researchers and raise their motivation. To take but one example, infant perception research draws huge numbers of workers to international conferences. There they can exhibit their latest research findings, argue with others, learn about the latest theoretical trends and place their own researches into a contemporary context. Outside the lecture theatres they also have fun.

Given that there is no general philosophical agreement among vision researchers on what needs to be explained about perception – conscious experience, neurophysiological mechanisms, and so on – it is not surprising that theories of visual perception have so far lacked the rigour and power of the great scientific theories. We should not be depressed by this fact. The brain is the most complex system in the known universe. It may never be fully understood.

The importance of methods and measurement

In editions 1 and 2 of this book, the first substantive chapter was devoted to the concept of the threshold. Students tended to find this chapter dry and over-technical; some academic colleagues wondered whether the theory of the threshold was a true psychological theory. On the second point, we believe that we were right. However, in response to students' complaints we have removed the chapter.

That said, we have found that students, particularly beginners, and also lay people who have read *Theories of Visual Perception*, often wish to know more about how certain phenomena in perception were discovered. This section says something about the relation between methods, discoveries, and the theories designed to explain the discoveries. We shall develop this point by describing a few examples.

In the eighteenth and nineteenth centuries, sensation was held to be instantaneous; that is to say, when a surface is touched and the touch is felt, there is no time lag between these two events. The reason for this belief is that for much of the period many thinking people believed that the nervous system was part of the soul. As such, it was part of God's creation and was therefore perfect and instantaneous. We should not, in retrospect, deride these beliefs; they were part of the culture of citizens educated in Christian communities.[2]

---

[2] Remember that, to this day, nearly 50% of American citizens reject the theory of evolution.

In the middle of the nineteenth century, the great scientist Helmholtz (of whom we shall learn more later) carried out a reaction-time experiment. He touched the foot of one of his observers and instructed him to push a key on feeling the touch. Many trials were carried out and the average time to react was recorded. Then Helmholtz repeated the experiment, this time touching the observer's thigh. On average, these reaction times were shorter. Helmholtz knew the distance between the foot and thigh sites. Thus, by subtracting the foot times from the thigh times, Helmholtz was able to arrive at an estimate of neural conduction time. This showed that the average conduction time was about 100 metres per second. Sensations are definitely not simultaneous with stimulus events.

This result changed many things. For example, it led to the measurement of reaction times under many different conditions (reaction times to pain, to heat, to cold, and so on). The development of measures of *choice* reaction times ('press this switch when a red stimulus appears, press that switch when you see a green stimulus') opened up all sort of possibilities of measuring human *decision times*: techniques that are used to this day in exploring subjective perceptual complexity and even the complexities of language. Reaction time research has been wonderfully fruitful. And many of the data gathered in this research have led to new and important theories. For example, reaction times to pricking pain are considerably shorter than reaction times to burning pain. This led eventually to the *gate control theory* of pain – a very important development.

Helmholtz was able to establish this breakthrough only because he was able to use a new precise timing device, the kymograph, on which were mounted two relay-driven pens. One was triggered by stimulus onset and the other marked the revolving paper drum when the observer responded. Simple reaction times (e.g., reacting to touch) average around 120 ms – less than one-fifth of a second. Until it was possible to measure with this degree of precision, human reaction times would have played no part in experimental psychology (see the Endnotes to this chapter).

Once accurate timers became available, it was possible to calibrate mechanical shutters. This led to the development of the tachistoscope, by means of which it was possible to present visual stimuli for very short periods of time (one-hundredth of a second, say). Immediately, it became possible to measure how much time it took to identify visual stimuli, to count clusters of dots and to read words. Use of modern electronic versions of the tachistoscope has allowed modern researchers to measure how long it takes to place a visual image on the retina (this proves to be nearly instantaneous) and how long it takes to feed information from the image back to the visual cortex (the rate here is about 10 items/s). Using even more sophisticated tachistoscopes, it has been found that an image can be 'wiped off' the retina by a second stimulus, provided the interval between the two exposures is sufficiently brief.

At start of the twentieth century anatomists and physiologists became interested in the functions carried out by different regions of the brain. Of necessity, many of these investigations were crude: a common technique, known as ablation, was to remove part of the brain of an experimental animal and then, after the animal recovered, it was tested to see which functions had been lost. Refinements in this technique eventually enabled researchers such as Sherrington and his colleagues (see, e.g., Creed, Denny-Brown, Sherrington, Eccles, & Liddell, 1932) to explore the working of the spinal cord in anaesthetized animals: it was in this manner that the first spinal reflexes were discovered.

A few decades after Sherrington's researches were published, the microelectrode was developed. A typical electrode is formed by pulling a red-hot hollow tube into a fine capillary. The tube is then filled with a conducting fluid. By carefully lowering the tip of such electrodes onto the visual cortex of living cats (and connecting the electrode to a powerful amplifier – another invaluable invention), Hubel and Wiesel (1962) were able to record the activity of individual cortical cells and eventually to show the functional architecture of this region of the brain. Hubel and Wiesel received the Nobel Prize for this work.

The availability of digital computers can be said to have revolutionized the study of visual perception. For example, prior to 1950 a typical perceptual experiment required the experimenter to present a stimulus to which the observer was required to respond. Then the next presentation was made and the process repeated until all the experimental trials had been completed. This simple process yielded masses of important data. However, as the present author can testify, running such experiments was a lengthy time-consuming process, and it was often necessary to test 10–20 volunteers in order to obtain reliable statistical data.

When digital computers became available, the situation changed dramatically. Simple experiments could be run in the absence of the researcher, the resulting data could be analysed immediately and the entire procedure was much more efficient and took less time. More importantly, the computer enabled a completely different type of experiment to be run. For example, in studies of eye movements, the subject could be wired up to a computer, then the computer recorded the current position of the eye and delivered a visual stimulus. This was merely an increase in efficiency. However, because of the speed of the computer, the visual display could be changed *during an eye movement*. This technique yielded important insights into what goes on when we make rapid saccadic eye movements – whether the eye can take in information during such movements, and so on.

As will be shown in Chapter 2, the perceptual capacities of newborn infants are of enormous theoretical interest. The challenge has always been how to communicate with these infants. A wonderfully simple technique depends on a phenomenon known as habituation. Show an infant a visual pattern and the infant will look at it. Keep on showing the same pattern and eventually the infant will pay it little or no attention (measured by filming the infant's direction of gaze). Now change the stimulus. If the infant starts to look at the new stimulus, this is a sign that the change has been detected. Careful manipulation of stimulus patterns in this research has led to some quite remarkable discoveries concerning what infants can see in the first days and weeks after birth.

Each of the technical discoveries outlined above has had a major impact on perceptual theory. Reaction-time data have led to theories of human decision making. Tachistoscopic experiments have taught us much about the nature of visual processing. Single cell recordings from the visual cortex have stimulated theories as to how the visual apparatus organizes inputs to create reliable representations of the external three-dimensional (3D) world. Eye movement studies continue to yield insights into the nature of language processing. There can be no doubt concerning the importance of new techniques in science.

Endnotes

- Some of the problems to be discussed in the remainder of this volume are philosophical ones. Readers who have not undertaken a formal study of philosophy will profit from reading sections in Gregory (1987) and the whole of Dennett (1991).

- In the following chapters there are numerous references to neurophysiological research and theory. Several good undergraduate texts contain excellent accounts of this material. Rosenzweig, Leiman, and Breedlove (1999) is a general account of biological psychology with good sections on the nervous system and perceptual mechanisms. The book is well illustrated and comes with a floppy disc suitable for Macintosh and PC computers.

- Although Helmholtz used human reaction time in an inspired manner to measure neural conduction speeds, he did not discover the reaction-time phenomenon. Many years earlier, astronomers found that different observers recorded different transit times when studying the motion of the planets. When a particular planet was known to be passing through a region of the sky, a telescope fitted with a graticule would be pointed towards the predicted position of the planet at a particular time. As the moving image of the planet was spotted in the telescope, the observer would wait until the planet met the edge of the graticule. As the planet's image crossed the central line, the astronomer would react by pressing a key and the transit time was noted. However, different observers recorded different transit times. They then visited each other's observatories and found the same effect. Eventually, these errors became known as 'personal equations' and were allowed for in estimating true transit times. What was happening, of course, was that different observers had different reaction times.

2 The Gestalt theory

The first general theory of perception to be discussed, Gestalt theory, represents a fascinating paradox. As a formal theory of perception, it can be said to have failed. However, it can also be asserted that the approach was, within limits, brilliantly successful and that it continues to exert a significant influence on the psychology of perception. We shall attempt to show how this paradoxical situation came about.

Gestalt theory is closely associated with the work of three men: Max Wertheimer (1880–1943), Wolfgang Köhler (1887–1967), and Kurt Koffka (1886–1941). There were (and are) other Gestalt psychologists, but these men pioneered the approach.

The remainder of this chapter will cover the following topics:

- The historical background to the movement.
- A general outline of the Gestalt approach.
- Köhler's brain model.
- A preliminary assessment of the Gestalt theory.
- Subsequent research on some of the Gestalt principles.

We shall begin by tracing some of the historical origins of this important movement.

Historical background

The benefit of hindsight allows us to discern some of the influences that made the emergence of the Gestalt theory almost inevitable.

Philosophy

When the philosopher Kant (1724–1804) published his *Critique of Pure Reason* in 1781, the book exerted a major impact on subsequent European philosophy. Nobody could do justice to this influential and difficult work in a few lines, but it is possible to take a single Kantian idea as an example of his thinking and show its potential importance for psychology. Imagine an object moving from left to right across the field of vision. The movement takes time and occurs through space. That we perceive the object's motion is something we take for granted; we can even guess about mechanisms that enable us to do this. But what about the framework within which the object moves? We cannot perceive space itself, there is literally nothing *to* perceive. Space is what the object moves through. Similarly, we cannot perceive time as such: it too is simply a framework within which events are ordered. But the perception we have had would clearly be impossible without our awareness of these frameworks. Where does the awareness come from? Kant's answer (stated here with some crudity) is that space and time are *a priori* intuitions. That is, they are 'givens', superimposed upon reality by our minds. Much of Kant's life was spent in justifying this claim and examining the consequences. For now we can say simply that, if Kant's claim is true, there is one consequence which is of great psychological as well as philosophical importance: perception must be in large part innately determined. Mind imposes a structure on the perceived world and this world is the only one we are capable of perceiving. This leads to nativism, which, as will be shown, became associated with the Gestalt approach. Gestalt psychology did not appear until 100 years after Kant's death, but it was clearly influenced by his philosophical investigations.

Nineteenth-century ideas

There were other influences at work in nineteenth-century Europe. Many people, as a result of Darwinism, had abandoned their religious beliefs. If God was no longer to be seen at the centre of things, who was? The answer was, man.[1] It is significant that the nineteenth century witnessed the great flowering of the romantic movement in literature and the arts. Common to much of the creative work associated with this movement is the struggle of the individual – a hero or heroine – against fate. Think for a moment about the great romantic poems of the period, the operas, emotions induced by the music of Beethoven (e.g., *The Eroica*: 'A heroic symphony . . .'), Wagner, and Tchaikovsky. The themes that emerge repeatedly concern the heroic struggles of individuals. This was still part of the Zeitgeist at the time of the emergence of the Gestalt movement, where the emphasis was on the dynamic role of the perceiver in making sense of the world.

The law of least action or the minimum principle

Ancient Greek geometers spent much effort in exploring what they considered to be ideal forms. They discovered that in the case of one such

---

[1] Historically, this was the common term.

ideal form, the circle, a fixed point on its circumference traces an interesting shape as the circle is rolled along a straight line. This shape is known as a *cycloid* (see Figure 2.1).

The publication of Newton's *Philosophiae Naturalis Principia Mathematica* in 1687 led to a renaissance in which mathematicians strove to discover the general laws governing the physical world. Nine years after Newton's work appeared, the Swiss mathematician Johann Bernoulli wrote a letter in which he described what is now known as 'the brachystochrone problem':

> For two given points *A* and *B* in a vertical plane, find a line connecting them on which a moveable point *M* descends from *A* to *B* under the influence of gravitation in the quickest possible way.
>
> (Quoted by Hildebrandt and Tromba, 1985)

The reader will realize from Figure 2.2 that there are two aspects to the brachystochrone problem: first, how to make the 'best' use of gravity; second, how to get *M* from left to right. Commonsensical but wrong solutions are shown in Figure 2.2a. The correct solution, shown in Figure 2.2b is a *cycloid*.

Interestingly, the seventeenth-century Dutch scientist, Huygens, realized that the inaccuracies of contemporary pendulum clocks were due to the fact that the time taken for a full period of a pendulum varies as a function of its

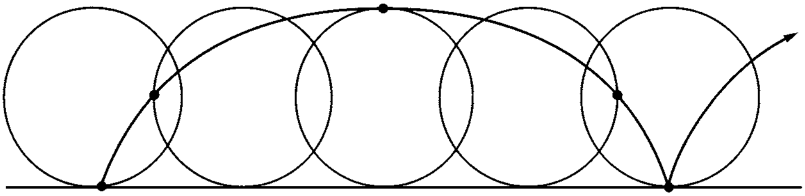


Figure 2.1 The shape traced by a point on a moving circle: a cycloid.

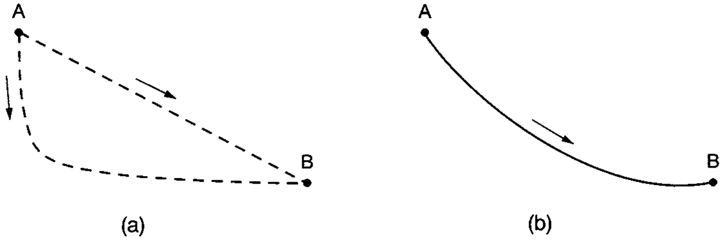


Figure 2.2 Bernoulli's problem. What is the most economical path for the movement from A to B under the influence of gravity? (a) Two incorrect solutions; (b) the correct solution – a cycloid.

amplitude – because the path of a simple pendulum describes an arc which is circular. In 1657 Huygens patented a solution to the problem – a way of causing the pendulum to swing in an optimal manner – which was to fit a collar around the flexible top of the pendulum, causing the swing to follow the path of . . . a cycloid.

In 1744, Bernoulli's pupil, the great mathematician Euler, applied the general principle of 'least action' to the motion of the planets around the sun. At the same time, Maupertius announced his law of least action: 'If there occurs some change in nature, the amount of action necessary for this change must be as small as possible'.

In the next century the existence of a general principle of minimum effort or action was demonstrated in a variety of phenomena, ranging from optics (light taking the shortest or quickest route) to the behaviour of chains suspended between supports (at equilibrium the chains adopt a curve known as the 'catenary'). Particularly interesting was the application of the principle of minimum effort to soap films. Such films, when stretched across frames, assume equilibrium states of minimum potential energy. As potential energy is proportional to area, it follows that the shapes assumed by the films are *minimal surfaces*: they represent the smallest areas capable of spanning the frames. As the examples above demonstrate, this work had interesting practical as well as theoretical implications. The tradition continues: the remarkable Olympic Stadium in Munich, built in 1976, has a roof built with high steel masts connected by steel ropes and covered by a transparent membrane. The complex shape of this roof is in fact a minimal surface; its design was arrived at by building models in which the masts were connected by films produced from soap solutions.[2] We shall say something about the latest developments concerning the minimum principle later in this chapter.

That was the historical background to Gestalt psychology. But the movement's beginnings were also a reaction against two contemporary approaches to psychology, structuralism and behaviourism.

Structuralism and behaviourism

Structuralism, which reached its peak between 1870 and 1910 with the work of Wundt in Germany and Titchener in America, was an attempt to explore the mind in a manner analogous to the chemical analysis of complex substances. Just as the chemist can consider compounds in terms of their basic chemical elements, Wundt, Titchener, and others believed that the laws of the mind would be revealed by careful study of its elements and their relationships. In this case, the 'mental elements' of the analysis were

---

[2] The ideas so far described in this short section are taken from Hildebrandt and Tromba's *Mathematics and Optimal Form* (1985). The interested reader is urged to consult this fascinating book.

sensations. On this view, any rich subjective experience is essentially a blend of simpler, more basic experiences or sensations, and the job of the psychologist is to list these. But this reduction to sensations is not easy. There is a constant tendency to commit the 'stimulus error', in which the source of a sensation is confused with the sensation itself. For example, one frequently says that one hears *something*, say an engine. But the engine is not itself a sensation. One should say that one has certain sensations *like* those normally arising when one is near an engine and then attempt to analyse these. The technique for avoiding the stimulus error and for correctly identifying the sensations one is having is a difficult one that must be learned and practised, usually by the investigator who is the subject in the research. The technique is known as trained *introspection* and was the basic source of data in the experiments of Wundt, Titchener, and others.

Reduced to its simplest formulation, structuralism leads to a view of perception in which the perceived world is a mosaic. Each stimulus element in a scene yields its own sensation and the totality of these sensations forms the percept. Stated baldly, it is obvious that such a scheme could not work. How to explain, for example, why things remain the same size as we move away from them if our perception is tied to particular sensations, in this case those arising from the shrinking retinal image? To avoid this trap, the major theorist Helmholtz (who was Wundt's contemporary and whose career overlapped with Titchener's) had been driven to an empiricism, which asserted that experience and memory must correct and enhance the momentary effects of stimulation (empiricism will be the subject of a later chapter). This in turn suggests that much of perceiving must be learned. While not denying the role of experience, the Gestalt theorists rejected both the mosaic view of perception and the emphasis on learning, both of which came under attack in their own writings.

Structuralist introspection as a method of studying perceptual phenomena is long dead and must strike the reader as somewhat unusual. It may help to finish with an example of the sort of thing that the Gestalt theorists attacked. Here is Titchener introspecting on the taste of two substances:

> Thus the 'taste' of lemonade is made up of a sweet taste, an acid taste, a scent (the fragrance of lemon), a sensation of temperature and a pricking (cutaneous) sensation. The 'taste' of limewater is made up of a weakly sweet taste, a sensation of nausea (organic sensation), a sensation of temperature and a biting (cutaneous) sensation.
>
> (Titchener, 1901)

Note that the phrase 'fragrance of lemon' implies that Titchener may not have fully succeeded in reducing the taste into its basic sensations: an analysis of the lemon fragrance is now required. The reader who would like to try this sort of thing is directed to the limewater problem; Titchener's use of the word 'nausea' is curiously apt.

The structuralists knew most of what there was to know about the experimental psychology of perception at the turn of the century. Their writings reveal close familiarity with the works of Helmholtz, for example. But introspection failed (and with it, this type of structuralism) for a number of reasons: trained observers frequently disagreed in their introspections; introspective data cannot be easily quantified; most importantly, many mental processes are simply not available to self-observation. In fact, the influence of structuralism was probably at an end by the time the first Gestalt discoveries were announced. However, as we shall see, the approach made a useful straw man for the Gestalt theorists, all of whom were gifted polemicists.

The second focus of the Gestalt theorists' attacks was behaviourism. Once again, it could be said that the Gestalt theorists exaggerated the influence of this movement, at least on contemporary work in perception (and it must be pointed out that the behaviourists were also hostile to structuralism). Nevertheless, it was true that behaviourists did attempt to explain behaviour in terms of a model derived from classical conditioning. This concentrated upon simple stimulus–response relationships, and tended to treat stimuli as essentially simple events confronting organisms. Further, the behaviourists had stated that the subject matter for psychology was objective behaviour, and only objective behaviour. Mental events, subjective experiences, had no place in this new, tough-minded scientific approach.

Gestalt theorists published lengthy rebuttals of the behaviourist case (see e.g., Köhler, 1947, Chapter 1). One of the most telling criticisms is that advanced by Köhler against the objectivity that behaviourists aspired to. Köhler argued that this was a chimera and that even in physics – which claims to deal with the objective world – the concepts and observations are never objective in the sense the behaviourists had assumed:

> How do I define my terms when I work as a physicist? Since my knowledge of physics consists entirely of concepts and observations contained in or derived from direct experience, all the terms which I use in this science must ultimately refer to the same source. If I try to define such terms, my definitions may, of course, refer to further concepts and terms. But the final steps in the process will always be: pointing towards the locus of certain experiences about which I am talking, and hints where to make certain observations. Even the most abstract concepts of physics, such as that of entropy, can have no meaning without a reference, indirect though it may be, to certain direct experiences.
>
> (Köhler, 1947)

It is worth noting that these remarks of Köhler's would have carried extra weight, as it was known that he had trained originally as a physicist.

To summarize, the Gestalt theorists were opposed to sensations as data and the accompanying mosaic view of perception, to crude atomism, to introspection as a method, and to the search for a bogus objectivity in psychology. These objections will acquire more force when we outline what Gestalt psychology offered in place of the approaches and assumptions to which it was opposed.

A general outline of the Gestalt approach

The start of the Gestalt movement

This is one of the best-known stories in the history of psychology. It should be said at the outset that two important Gestalt principles had been published prior to the formation of a separate Gestalt school of thought. Ehrenfels (1890) had drawn attention to the fact that many groups of stimuli acquire a pattern quality that differs from the parts when seen in isolation: a tune is more than the sum of its notes; in a square something emerges which has a quality not present in a random assembly of the component lines – the 'squareness'. Ehrenfels named this emergent property, '*Gestaltqualität*' (form-quality), a name that was adopted by the Gestalt movement. A second precursor of the Gestalt movement was Rubin (1915), who published an important paper on the distinction between figure and ground in perception, a distinction that later found an important place in Gestalt thinking.

In the summer of 1910, the person who can be said to be the true founder of Gestalt psychology, Max Wertheimer, broke a journey in order to buy a toy stroboscope. He then carried out some investigations of the illusory movement that such devices can create. If one exposes two stimuli alternately in rapid succession, then a number of strange things can happen, depending on the exposure times, the rate of alternation, and so on. At low rates of alternation, two separate stimuli are seen; at higher rates one sees a displacement of a stimulus from one position to the other (this can be seen at British Rail unmanned crossings, where pairs of red warning lights flash alternately): this is stroboscopic movement. But there is an optimum rate at which what is seen is not a moving stimulus, but simply movement *per se*. Obviously, this movement cannot be explained in terms of the behaviour of either of the two stimuli – each simply appears and disappears at its own location. The experience of pure movement, which Wertheimer later called phi-movement, arises as the result of temporal and spatial *relationships* between stimuli: something new has arisen which differs from (is over and above) the sum of the parts acting in isolation – it has *Gestaltqualität*.

Wertheimer continued to work on the phi phenomenon at Frankfurt University, using as subjects two young psychologists, Wolfgang Köhler and Kurt Koffka. This trio were to create a new approach to the study of perception and a major theory – the Gestalt theory. It is perhaps unfair to single out one member of the movement for a biographical sketch. However, a few remarks concerning Köhler's career will convey something of the academic life of the period. Köhler was educated in three German universities and was at the Psychological Institute in Frankfurt when Wertheimer started his work on the phi-phenomenon. Köhler worked from 1913 to 1920 in Tenerife. He was really there as a German agent, sent by his government to spy on allied shipping. His cover was the activity of studying problem solving in chimpanzees. Later, after directing the Psychological Institute in Berlin, he fled Nazi Germany in 1935 and spent the rest of his academic career in the USA. His last post was at Dartmouth College.

Phenomenology

There is a special way of looking at the world. To experience it, follow this simple procedure: take a piece of paper and punch out a small hole in the middle about half a centimetre in diameter. Examine any nearby surface and note its colour. Look again, this time through the hole, with the paper held about six inches from the face. Two things may become apparent. First, the colour seen through the hole (which has the fancier name, 'reduction screen') no longer appears to belong to the surface; it seems to float as a film just behind the hole. Second, colours may appear different from those seen when looking normally: for example, someone sitting by a wall may have a portion of their face tinted in the wall's colour; grey shadows may now appear coloured.

Using a pencil in front of one eye as a referent (a trick commonly used by artists), one can quickly come to see that objects subtend smaller visual angles as they recede from us. And, once it is pointed out to us, we can experience this troubling fact: the nose is always visible in our field of view – if we choose to notice it.

The procedures above are not simply tricks. It is a fact that careful analysis of what is to be seen when we look in a special way differs from what we normally experience. Introspectionists believed that demonstrations such as the reduction screen reveal the raw material of perception, namely sensations. And it is undeniably true that retinal images of objects do in fact shrink with distance, and that we can become sensitive to these changes: artists must.

The question the Gestalt theorists raised was, which of these two modes of perceiving should be explained in perceptual theory? Their answer was unhesitating and forceful: everyday experience. To this end Koffka asked what has become the most famous question in the history of perception: 'Why do things look as they do?' In other words, what must be explained by perceptual theories is the stability and coherence of the world of everyday experience, the world in which surface colours are stable under different illuminants and familiar things do not change size as they recede. This is a world of objects, not sensations, and the proper approach to this world is that of the phenomenologist.

> There seems to be a single starting point for psychology, exactly as for all the other sciences: the world as we find it, naively and uncritically.
> (Köhler, 1947, opening paragraph)

The decision to try to understand the world of the unselfconscious perceiver shaped Gestalt research and led to the distinctive *style* of the movement. Gestalt workers concentrated mainly upon strong effects in perception, a legitimate approach, but they went further: whenever possible their readers are offered, not a table of experimental results, but a compelling illustration. The emphasis is upon experience rather than data. The reader is to be convinced, not by the results of some experiment, but by what he or she actually sees. The unusual power and clarity of Gestalt writings owes much to this tactic.

Perception as a dynamic, organized process

Whenever we open our eyes we see, not sensations of light, but objects and surfaces. There is a tendency (most easily noticed in vision) to organize our percepts in a certain manner during all perceiving: we effortlessly distinguish between the *figure* in a field of view and the *ground* against which it is seen. The figure–ground distinction is highly important evidence for the dynamic character of perception. Figures tend to be complete, coherent and in front of ground, which is seen as less distinct, is attended to less readily, and is often seen as floating behind the figure. When figure and ground share a contour (as they commonly do), then the contour is usually seen as belonging to the figure.

In Figure 2.3 the immediate organization leads us to see a black triangle (the figure) in front of a white ground. But the printed page permits trickery. The white disc on the triangle, is it a figure? – in which case it will be seen as over the black (which is now ground) or is it an aperture? – in which case we appear to be looking through the triangle at the white

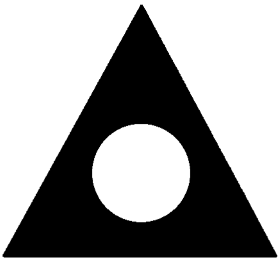


Figure 2.3 Ambiguous figure–ground relationships. Is the white disc superimposed on the triangle, or is it a hole through which the underlying ground can be seen?

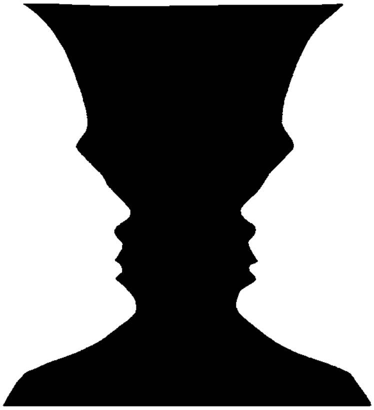


Figure 2.4 Figure–ground reversal: the face–vase illusion.

ground on which the original triangle is superimposed. Notice the subtle change in the status of these figure–ground relationships when we change our attention in this way. Figure 2.4 shows how figure–ground relationships can be made entirely ambiguous: which is the figure in this case, the profiles or the vase?

Figure–ground separation occurs in all sensory modalities, for example when we abstract the voice of a speaker from the background sounds of a noisy party, or when we feel an insect crawling over our skin. And it seems that we do not have to learn how to achieve this valuable economy in perceiving. When people recover their sight after many years of blindness they commonly experience many difficulties in seeing the world as it is. However, almost without exception, the case reports say that figure–ground separation is achieved from the outset. Are we built to see in this way?

So powerful is the tendency to organize vision into figure and ground that we take it very much for granted – hence the Gestalt theorists' use of ambiguous material such as Figure 2.4, which is intended to shake us out of our normal habits. The magnitude of the figure–ground achievement becomes apparent to those attempting to make machines that can perceive. How could a computer be programmed to ignore everything but the people in a complex scene? What rules would enable it to record only the left-hand performance of a jazz pianist?

The laws of grouping

In one of the early discoveries in Gestalt psychology, Wertheimer (1912) demonstrated several principles by which groups of stimuli organize themselves in perception. Looking at the arrays illustrated in Figure 2.5 reveals a spontaneous tendency to organize the stimuli into wholes or *Gestalten*. For example, stimuli that are adjacent tend to be grouped together: in

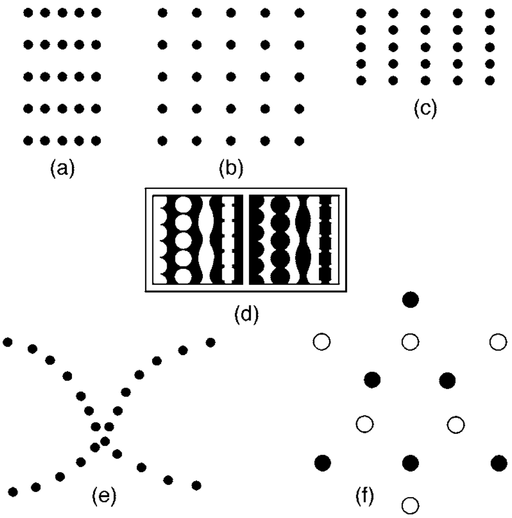


Figure 2.5 Some of Wertheimer's laws of grouping. (a) Proximity induces grouping by rows; (b) proximity is equal and there is no dominant direction of grouping; (c) proximity induces grouping by columns; (d) grouping by symmetry; (e) grouping by continuation; (f) grouping by similarity.

Figure 2.5a the stimuli could be seen as unconnected, as rows, or as columns. But the adjacency or proximity principle guarantees that we see them as paired columns. The figure illustrates some other laws of grouping, such as good continuation, similarity, and closure (the tendency to see a completed figure whenever possible). If a subset of the stimuli in Figure 2.5 were to move in the same direction, then this movement would cause them to separate phenomenally and take on organized figural properties, illustrating the law of common fate.

These spontaneous groupings in perception are fascinating and reliable phenomena and were still being researched 80 years after Wertheimer's demonstrations (see e.g., Restle, 1979). It is difficult, having experienced such effects, to return to any view of perception that ignores its dynamic aspects. Note once again the power of demonstrating rather than describing phenomena.

Goodness or Prägnanz

The Gestalt theorists concluded that there must be a general underlying principle behind the numerous examples of organization that they discovered. It was as though perception tended, wherever possible, towards simplicity, symmetry, and wholeness, a tendency summarized by the German word, *Prägnanz*. As applied to perceptual phenomena, the concept of Prägnanz is in fact a rather complex one.

In modern German, the word 'Prägnanz' can mean clear-cut, concise or succinct. But as Arnheim (1987) states, Prägnanz can imply not only a tendency toward regularity and symmetry – the cleansing of the stimulus of distracting detail – but also the intensification of characteristics. For example, when we suddenly see a face in the amorphous configuration of a cloud or a dying fire, this is change towards perceptual simplicity. However, once the face appears, the details become emphatic. If anything, this is a tension-enhancing, rather than a tension-reducing, process.[3]

We now call the process of seeing novel similarities 'lateral thinking'. Remembering that Köhler had trained as a physicist helps us to understand the next stage in the development of Gestalt theory. Where else do we find processes that tend towards simplicity? The answer, as was demonstrated earlier, is in the physical world. Perception appears to be analogous to certain processes that we can observe in nature. It is an exciting step to wonder whether essentially similar physical forces are the cause of Prägnanz in perception. Later we shall show that this is the conclusion that Köhler eventually arrived at.

Wholes and parts

The claim that in perception the whole is different from the sum of its parts acting in isolation is one of the most important tenets of Gestalt theory. This simple idea, elegantly illustrated in numerous demonstrations, has great significance for perceptual theory. If correct, it rules out the possibility of developing adequate theories of perception that treat stimuli as isolatable events.

As was stated earlier, Ehrenfels (1890) introduced the concept of *Gestaltqualität* prior to the emergence of Gestalt psychology. The importance of this concept cannot be exaggerated and it was elevated to a major principle in the Gestalt theory. When we hear a tune, the experience of the tune itself (the *Gestaltqualität*) is something more than the aggregate of the notes. It is not reducible to individual notes and is not an adding together of simple sensations. For example, the last three notes of the British national anthem are the same as the first three notes of *Three Blind Mice*, but how many people who know both tunes well have ever realized this? The notes do not sound the same because they are not in the same context. And there is an interesting paradox here: although the notes form the context, the context shapes the notes. A similar effect can be seen in Figure 2.6. In this illustration the top row is a set of simple shapes, hardly describable as an organized figure. In the lower row the shapes have been grouped in a certain way. The shapes are now organized into a face. Note that the face would not

---

[3] The author is indebted to Professor Lester Krueger of Ohio State University for advice about the subtle meanings associated with the term 'Prägnanz'.

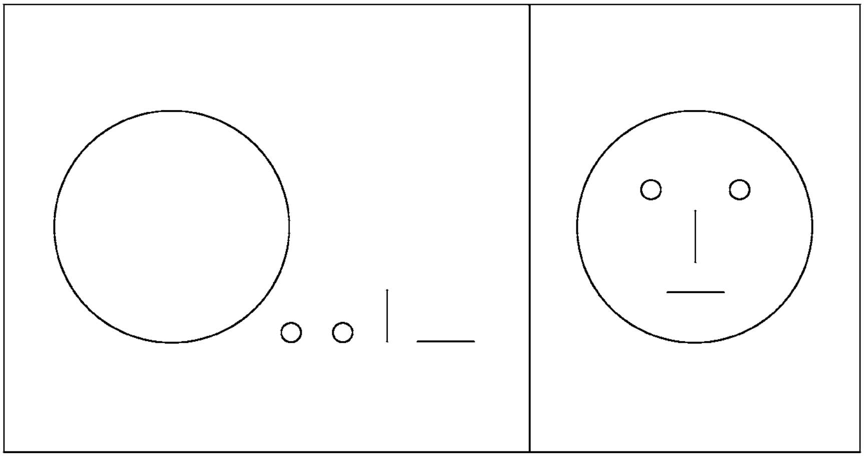


Figure 2.6 The mutual interaction of parts and wholes: the simple shapes, when assembled in a certain manner become organized into a recognizable pattern – the face. But within the face the parts acquire new meaning.

have emerged without the shapes, but now the shapes themselves are seen differently: circles become eyes; a line becomes a mouth, and so on. The parts form the whole, but the whole changes the parts. When a tune is transposed an octave, or played in a different key, we recognize it as the same tune even though each individual note is different. Because the *relationship* between the notes is the same, they exhibit the same Gestaltqualität.

Historically, the most influential demonstration of part–whole interactions was Wertheimer's use of the phi phenomenon, which has been described earlier. Phi movement is something new, something not predictable from the behaviour of each light in isolation, but emerging as a function of the spatial and temporal *relationships* between the lights.

In Figure 2.7 the triangles have been formed from different elements, and yet 'triangularity' is evident in each display. None of the parts in isolation possesses triangularity; this emerges only in relationships. The Müller–Lyer illusion in Figure 2.8 is another example of this important point: the shaft lines are objectively the same length, but their relationship with the arrows creates an illusion, an illusion which could not have been predicted from knowledge of the individual components.

The constancies

The tendency for perception to be veridical, summarized by the term 'perceptual constancy', was seized upon by Gestalt theorists. When objects recede they commonly do not shrink; white paper in shadow does not look greyer; objects remain the same colour despite changes in illumination; shapes do not change when seen from new positions.

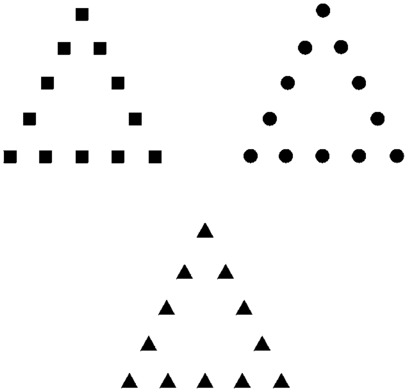


Figure 2.7 Triangularity Gestalten.

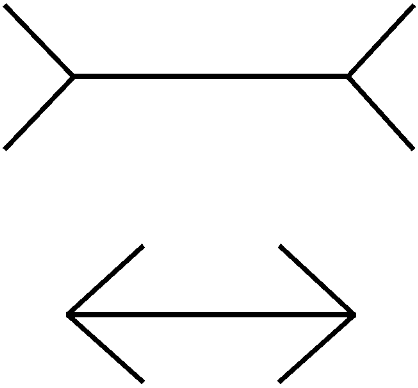


Figure 2.8 The Müller–Lyer illusion.

All these are examples of perception going beyond the local effects of isolated stimuli. In these cases, Gestalt theorists likened the environmental context in which stimuli are lodged to a dynamic '*field*', a term synonymous with that currently being developed by the physicists of the period:

> The constancy of brightness, for instance, depends on the relation of the illumination and brightness of the surrounding field to the brightness of the object under observation.
>
> (Köhler, 1947)

We shall return to the idea of a field later in this chapter.

The main perceptual phenomena which shaped the Gestalt theory have now been outlined. During the history of the Gestalt movement the work was extended to other areas. It was found, for example, that when a chicken is trained to peck at the darker of two greys and this is now paired with an even darker grey, it is to the latter which the animal now responds – suggesting that the original learning involved a *relationship* rather than an absolute stimulus value. Monkeys solving problems which are more closely related to their natural lives than the laboratory mazes used by the early behaviourists do not engage in constant trial-and-error, but show periods of inactivity followed by sudden solutions to the problems. This suddenness following a latent period is characteristic of much human experience also. For example, when we suddenly see a face in a fire which we have been staring at for a long time, or when we equally suddenly solve a crossword clue. In a very different context, Gestalt theory has been applied to artistic phenomena (Arnheim, 1949, 1956, 1969). Some have tried to relate it to psychological therapies. However, it is in the field of perception that the theory has had its major impact.

It must be remembered that the ideas we have outlined above became known partly as a result of the flair and conviction with which they were announced. As has been said, the Gestalt phenomena were often demonstrated on the printed page. The Gestalt theorists were good writers and enjoyed polemics. The success of the movement is hardly surprising, and readers are urged to consult some of the original Gestalt writings cited in the endnotes to this chapter to experience more directly the power of this approach to perception.

Köhler's brain model

So far, we have used the term 'theory' very loosely in describing the Gestalt theorists' work. We have used it to describe a movement in the history of perception, some beliefs about the nature of perceivers and the ways in which their abilities should be studied, and a set of laws describing the behaviour of stimuli during various interactions. What is missing from all this has been an account of the Gestalt theorists' views as to *why* perception is as they claimed. This is an appropriate place to turn to the explanations that Gestalt theory advanced to explain the Gestalt laws.

If the word 'failure' can be applied to any part of Gestalt psychology, it is here. Historically, the neural explanation of Gestalt phenomena has never achieved the status and acceptance afforded to the empirical parts of the work, and it is instructive to consider why this should be so. Why is perception dynamic? What causes the degree of organization that we have described? How shall we predict the behaviour of stimuli in new situations – how do we know what something will look like? A set of descriptions cannot answer these questions. What are required are explanations. Not surprisingly, the Gestalt theorists, particularly Köhler, went to considerable lengths to meet this challenge.

With hindsight we can consider the problem facing Gestalt theorists as a choice between three alternative ways of explaining perceptual phenomena:

introspection, physical Gestalten, or physiological mechanisms within the central nervous system.

An introspectionist approach to the explanation of Gestalt phenomena would have been essentially psychological or mental in flavour. However, as we have seen, the hostility of the Gestalt movement to introspection rules out any explanation of this type.

Physical Gestalten and the minimum principle

Köhler considered the possibility that perceptual Gestalten were manifestations of a wider set of phenomena that included physical Gestalten. The reader will remember from an earlier section that studies of physical phenomena over the centuries had accumulated a mass of evidence to support a general principle of least effort or least energy – the *minimum principle*. It is hardly surprising that Köhler, as a trained physicist, should have been familiar with the minimum principle and the resulting tendency of physical systems to settle into equilibria involving minimum energy, minimum surfaces, and so on. Köhler's detailed knowledge of magnetic and electrical *fields* led him to suppose that if such fields behave in dynamic ways, exhibiting tendencies to closure, balance and Prägnanz, then perception might obey the same laws – given that it arises from the action of neural (physical) systems in the brain. In this way, Köhler attempted to attain 'scientific citizenship' (Petermann, 1932) for the concept of the Gestalt. Thus, in his work *Physical Gestalten* (1920), Köhler states that when two electrolytic solutions are in osmotic contact, the electrical potential that arises is a new property of the system as a whole: '. . . the communicating system of solutions has Gestalt characteristics'. It should be obvious what Köhler was trying to do in his discussions of these physical analogues.

Unfortunately, this approach cannot be fully sustained. It is possible to collect many instances of Gestalt-like phenomena in the physical sciences, but it is only too obvious that there are many situations in which assemblies of things, including chemicals, do not show Gestalt effects. Is a pile of coal a Gestalt? What about a mixture of salt and sand – where are the Gestalt interactions here? It seems unlikely that physical Gestalten have sufficient generality or relevance to permit extrapolations to psychological phenomena; they cannot carry the theoretical weight.

Even if the idea of physical Gestalten had seemed more plausible as the basis of a theory of perception, this would not have solved all the problems facing the Gestalt theorists. When we attempt to analyse a particular pattern, what can we say about its components? The Gestalt movement (at least after 1929) opposed any form of reductionism, believing that theoretical explanations in psychology should be 'from above' rather than 'from below'. More seriously, we find in Köhler (1925) the statement that in completion phenomena, which are excellent examples of the dynamic aspects of perception, '. . . a part will suggest a whole only if it is a genuine part'.

However, as one commentator has remarked, '. . . it is difficult to see how [this definition] can finally avoid the tautology that what produces a genuine whole is a genuine part' (Staniland, 1966).

The Gestalt theorists found themselves forced into an even more extreme position than this. Replying to a critic, Koffka was driven to say:

> . . . in characterising a real object as a stimulus we do not refer to an absolute property of that object, in and by itself, *but only to the object's relationship* to a living organism.
>
> (Koffka's reply to Benussi, 1915, emphasis added)

Later, Koffka adds:

> Hence even if there were no physical Gestalten, there might nevertheless be stimuli for Gestalt presentations.
>
> (Koffka, 1915)

As Staniland (1966) goes on to comment:

> If the perceptual experience cannot be inferred from the physical data and the stimulus data are not available to introspection, the only correlation left is with the processes of the central nervous system, and it was towards this that Gestalt theory deeply committed itself.

And in doing so the Gestalt theorists lost their chance of bringing about a permanent change in the ways in which psychologists approached the problem of explaining perceptual phenomena.

Isomorphism

Köhler attempted to lay the theoretical foundations for an adequate account of Gestalt phenomena (see e.g., Köhler, 1940, 1947). His writings placed a major emphasis upon physiological/neural mechanisms as the required level of explanation. To this end he announced a 'general leading principle', that of *psychophysical isomorphism*, in which it was assumed that there is a correlation ('coordination') between psychological experiences and physiological events in the central nervous system:

> Experienced order in space is always structurally identical with a functional order in the distribution of underlying brain processes.
>
> (Köhler, 1947)

For example, if the organization of a visual display leads one to group stimuli into, say, a triangle, then the stability and Gestalt character of the triangle is due to underlying processes in the visual cortex. These preserve the essential relationship between the components of the figure, in other words, their triangularity. If the triangle is formed from three sets of dots, then the underlying processes must preserve (a) the ordering of proximities and (b) the angular relations between the sides thus formed. Similar principles relate temporal ordering of experience to temporal sequences of brain processes.

The representation, it must be stressed, is described as *topological* rather than *topographical*. Just as the London Underground map indicates the correct sequence of stations, but would be of little use when navigating one's way through the streets above, so we must not expect that when someone reports that an array has become organized into a triangle, an actual triangle of neural responses has formed in the visual cortex. What have come into existence are neural processes underlying the spatial essence of the organized figure experienced as a triangle. It is important to recognize that Köhler did not suggest that there were pictures in the head, although many commentators have falsely accused him of this (see Henle, 1984, for a review of the many erroneous interpretations of Köhler's position). He knew that this merely displaces the problem (who, or rather what, perceives the pictures?). Gestalt isomorphism was that existing between organized experience and processes in the brain.

Köhler had been struck by the tendency of some stimulus patterns to reverse after a period of prolonged inspection. It was as if perceiving involved a process in the brain that caused its own termination. What sort of process could this be? Köhler knew about the behaviour of chemicals in solutions. Most readers will be aware that many substances decompose in solution into particles having (opposing) electrical charges. Water itself forms positively charged hydrogen ions and negatively charged hydroxyl ions: H^+ and OH^-. Common salt (NaCl) forms Na^+ ions, which have lost an electron, and Cl^- ions, which carry an extra one. Hydrochloric acid forms H^+ and Cl^- ions. Collecting differently charged ions at two spatially separated sites forms the basis of the electrical cell, which, when short-circuited, will cause a current flow. This is the key to the Gestalt brain model. Köhler therefore speculated that the following chain of events follows visual stimulation (we shall modify one of his own examples).

Consider a stimulus array comprising a light disc on a darker background. The disc is seen as an organized figure against a ground. The neural processes associated with the perception of the disc and the background terminate in the visual cortex. The final neurons in the causal chain between the retina and the visual cortex discharge chemicals into the fluid medium surrounding them and ionic decomposition takes place. The stronger discharges associated with the disc lead to higher ion concentrations at certain sites compared with those induced by the darker ground. But both figure- and ground-induced discharges are part of the larger liquid environment surrounding the visual cortex. Thus, electromotive forces will arise between the figure, the darker ground, and the internal environment. These forces will maintain a current, the intensity of which will depend upon the intensity of the original visual stimulus. Further, *the currents in the visual cortex will come under the influence of physical laws*. For example, they will distribute themselves spatially according to the laws of electrostatic vectors and therefore tend towards Prägnanz (Figure 2.9 illustrates this type of analysis). Thus, the dynamic tendencies which we can observe inside ourselves when perceiving reflect the influence of physical forces in our brains. Here we see the influence of the minimum principle (described earlier in this chapter) on Köhler's ideas: dynamic physical systems tend to stability and minimum work; perception tends towards Prägnanz; the brain is a dynamic physical system.

The reader should now have a good impression of the style of Köhler's thinking. It is clear that his brain model is quite a gross one, in the sense that it involves large areas of the visual cortex. (Towards the end of his career, Köhler was starting to speculate that the fields which he had postulated to account for the Gestalt-nature of perception might be smaller in scale, involving activity around single synapses: see Henle, 1984.)

The implications of Köhler's model are as follows. First, context effects are explained because the electrical processes in the brain are not local and discrete but behave as fields. Thus, the impact of a stimulus is determined in part by the nature of the surrounding array – a fundamental Gestalt principle. Second, the effects of stimulation can outlast a stimulus. Köhler does not claim that we can be aware of neural processes directly, but that we can adopt certain procedures that give us a clue to their nature. For example, prolonged fixation of patterns may give rise to *after-effects*, in which it can

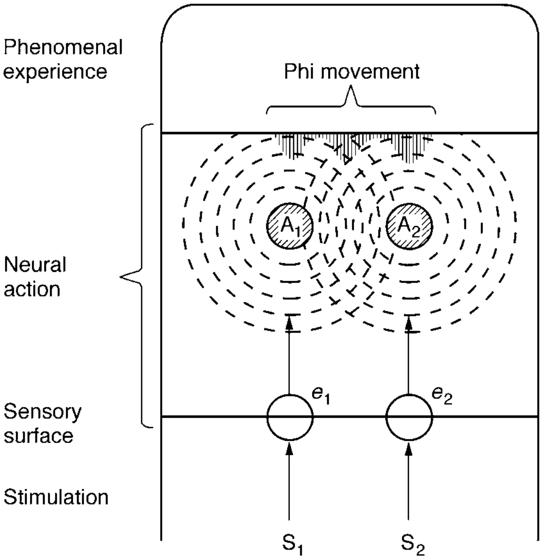


Figure 2.9 Köhler's concept of field forces. S_1 and S_2 are points in the visual cortex from which the induced electrical fields spread. The overlapping of the two fields A_1 and A_2 yields a unitary percept: in this case, phi movement.

be seen that one's perception is changed for relatively long periods. Thus, when one fixates a rotating spiral for several minutes, the after-effect (which is an apparent rotation of the spiral in the opposite direction) may last for hours or even days. Third, the behaviour of the fields in the visual cortex explains the tendency of perception towards Prägnanz: it is because the underlying electrical distributions follow minimal principles, and therefore tend towards balance and symmetry, that perceptual experience does the same.

Finally, the dynamic processes in the visual cortex have an existence of their own – they are physical events. If electrical charges can cross gaps and distribute themselves around resistance networks in a dynamic, holistic manner, we should not be surprised to discover that perception can fill gaps or show field-like effects: the two sets of phenomena are directly related.

The temptation to look back disparagingly at Köhler's psychoneural model should be resisted. Köhler was searching for an explanation of perceptual phenomena in terms of neural activity. There are still those, a quarter of a century after Köhler's death, who believe that the ultimate explanation of all psychological phenomena will be written in the language of physiology. His attempt was not absurd. Remember, too, that when Köhler first described his model, knowledge of the workings of the brain was sketchy, to say the least. It has been said that early neurophysiological attempts to probe the brain were like trying to understand people in the street from the top of a skyscraper, armed only with a giant needle. Köhler had to make the most of the knowledge available at that time. He wanted his model to be scientific. What could be more reasonable than to link it to some of the best science of his day, namely that associated with physics and chemistry?

To end this section on the Gestalt explanation of perceptual phenomena, it must be stressed that it stands apart from the main contributions of the Gestalt theorists. Other Gestalt psychologists were less concerned with Köhler's psychoneural model. Koffka, for example, was more content to describe and discuss Gestalt phenomena than to try to find a sound physiological explanation of them (see Koffka, 1924, 1935). To repeat a point made earlier, the wider meaning of Gestalt theory includes the phenomena described by Gestalt researchers and the psychological laws that they advanced. All this needs to be assessed. But for many who have supported the Gestalt approach over the years, the truth about the wider aspects of Gestalt psychology does not depend upon the correctness or otherwise of Köhler's brain model. We must respect this distinction when examining the Gestalt contribution to perception.

A preliminary assessment of the Gestalt theory

This assessment will attempt to look at the more general Gestalt contribution to perception (which will now be referred to as the Gestalt theory), rather than limiting itself to the brain model, although this will be mentioned.

We shall begin by describing some work that has challenged Gestalt ideas, to varying degrees. Then a brief selection of more recent work will be presented to show some of the directions in which research in this general area has gone since the earlier days of the movement.

As has been shown, the Gestalt theorists held that there are certain phenomena that reveal the basic laws of perception, that perceptual processes are dynamic, rather than passive, and that the perceptual world is organized into patterns or configurations rather than a mosaic of sensations. They argued for a phenomenological rather than an introspective approach to perception, and preferred strong demonstrations to statistical descriptions. Their explanation of perceptual and related phenomena took the form of hypothetical brain processes that were part of a psychoneural isomorphism, an explanation that is inherently nativist in its implications concerning the origins of perception in the individual perceiver. We shall now look at some of these topics.

The brain model

Earlier in this chapter an attempt was made to give as sympathetic an account as possible of this aspect of the Gestalt theory. Even so, the reader must have felt that we were drifting into science fiction. In truth, apart from Köhler, few have taken this account of perception very seriously. It is now possible to assert that it is probably wrong. (However, see Henle, 1984, for a vigorous defence of the general principle of isomorphism.)

First, while it is true that the disturbance of activity in the central nervous system following sensory stimulation often outlasts that stimulation, it is not true that this involves very large fields or areas of the brain (Sperry, Miner, & Meyers, 1955). Second, and more seriously, experiments have cast grave doubts on the existence in the brain of anything like the direct currents proposed by Köhler. Placing connected metal pins in the visual cortex of experimental animals should surely short-circuit such currents if they exist; the insertion of insulating mica plates into the same regions of the brain would be expected to block the spread of electricity. Both of these experiments have been performed and in neither case was the visual performance of the subjects seriously disrupted (Sperry & Miner, 1955; Sperry, Miner, & Meyers, 1955). We must conclude that this important part of the Gestalt theory, in fact the basic explanation of Gestalt phenomena, is likely to be incorrect (once again, see Henle's 1984 paper for a criticism of these direct attacks on Köhler's model of cortical functioning).

The inadequacy of two-dimensional displays

Because they wished to convince their readers through dramatic illustrations, and (presumably) because drawings are simpler to make than three-dimensional objects, the Gestalt theorists obtained many of their effects from flat patterns. Under these conditions, Wertheimer's laws of organization have not been seriously challenged: none has been shown to be actually wrong, neither is the general phenomenon of figure–ground organization recognized as being other than very important. However, the two-dimensional drawings that have been most commonly used to investigate these phenomena are not characteristic of all our daily experience: they are not what our eyes evolved to see. It is not surprising, therefore, that when three-dimensional arrays have been studied, the results have sometimes cast doubt upon the adequacy of the Gestalt laws (Kaufman, 1974, provides an excellent introduction to this later work). Further, there is at least one rival to the Gestalt view of figure–ground phenomena that places much more emphasis upon learned rather than innate factors (Hochberg, 1971).

In the next chapter, and elsewhere in this book, the concept of *ecological validity* will be discussed. For now, it suffices to define the phrase as meaning the naturalness of stimuli, how representative they are of the objects and events that organisms must deal with in order to survive. It must be said that many of the displays the Gestalt theorists used in their work had very low ecological validity. This does not prove that Gestalt generalizations are invalid or that Gestalt claims about the laws of perception are seriously wrong; simply, that their choice of stimuli was often unfortunate and should have made them cautious about over-generalizing their findings. This criticism can be extended to the Gestalt work on illusions, which are fascinating and reliable phenomena, but how often do we experience strong illusions in everyday life?

Stimulus ratios in perception

The Gestalt movement was correct in stressing the role played by ratios between stimuli as determinants of how things will appear. Paper always looks white and coal black across a wide range of normal light intensities. What is the basis of this veridical perception? One answer adopted by the Gestalt theorists was that in this case the paper would always reflect *relatively* more light than coal, no matter what the level of illumination. Similar arguments were advanced to explain other forms of perceptual constancy. Stimulus ratios appear to be important in perceiving.

Let us develop this further by adding some numbers to a possible case involving brightness (or lightness) constancy. Suppose that one surface reflects 25% of light falling onto it, a second surface 50%, and that light of, say, 200 units intensity illuminates both surfaces. They will reflect 50 and 100 units, respectively. Now double the illumination strength to 400 units: the surfaces will now reflect 100 and 200 units to the eye – although the illumination has doubled, the *ratio* of reflected light to the eye has not changed. This is the basis of the classic explanation of constancy that the Gestalt psychologists adopted to explain the unchanging lightness or brightness of surfaces under conditions of changing illumination.

Generally, there will be few real-life situations that differ markedly from the hypothetical example above. Reflectance is a property of surfaces: no change in illumination will alter ratios of reflection. The Gestalt explanation of brightness constancy seems secure. There is, however, one situation in which ratios of reflected light can be altered; this is when additional light is added, not to the surfaces but to stimulus energy on its way to the eye. This can happen when we look at things through a glass surface when that surface is reflecting additional light from other sources. One example is looking through the reflected glare of a shop window.

Gilchrist and Jacobsen (1983) noticed something that any Gestalt psychologist could have seen: things do not seem markedly different when we see them through reflections. These researchers describe an elegant experiment in which scenes were viewed through a sloping glass surface onto which additional light could be projected, thus altering the ratios of lights coming from objects in the scenes. This additional light is known as a 'veiling luminance'. To illustrate this point, consider again the second numerical illustration above: adding 100 units to the light that has already been reflected from the two surfaces will produce intensities at the eye of 200 and 300 units. The ratio has now changed from 1:2 to 2:3. So when we look *through* a reflection, the ratio basis of brightness or lightness constancy has gone. But Gilchrist and Jacobsen's experiment showed that, provided the scene is a real one (i.e., it has three-dimensional objects in it) *perception is veridical through a veiling luminance*. In other words, lightness constancy remains even though the ratios of reflected lights from the surfaces have changed. Thus, while the Gestalt emphasis on ratios is probably close to the truth, this cannot be the whole story. We still do not know exactly *how* perceivers use ratios to achieve constancy in perception, or how they cope when the ratios are corrupted, as in the above example. Similar conclusions have been arrived at by those who have examined the role of ratios of wavelengths of light in the perception of colour (see e.g., Land, 1985). Generally, we can suggest that most workers in the field of visual perception accept the importance of stimulus ratios, particularly in the perceptual constancies. However, there are cases in which it seems that there must be other effects at work in everyday perception, some of which are dependent upon more cognitive, knowledge-based processes involving judgement, familiarity with objects and surfaces, 'allowing for illumination', and so on (see Rock, 1995, for a lucid discussion of some of the theoretical problems in this area).

Stimulus ratios in learning

The discovery that an animal trained to go to the darker of two grey stimuli will subsequently transfer this learning to an even darker grey was held to be an important extension of a Gestalt principle from human perception to animal behaviour. It was claimed that the demonstration was particularly embarrassing to stimulus–response theories of animal discrimination learning. It suffices to say that, in fact, stimulus–response theories can be made to account not only for this result but also for those instances when the phenomenon does not occur (e.g., when the differences between the various stimuli are very large). This was demonstrated by Spence (1956).

Nativism

A few preliminary remarks are needed here. Earlier in this chapter we outlined Kant's views on the origins of how we come to perceive the world. Kant's approach can be labelled a 'rationalist' one. This was in marked contrast to the views of the English philosopher John Locke (1632–1704), who held that at birth the mind is a blank slate ready to be written upon by experience – a philosophical position known as empiricism. Both these philosophers were concerned with the problem of knowledge. Later, when psychologists became concerned with a broader but related set of problems, the debate continued as the nature–nurture or the innate–acquired controversy.

There have been many who consider this controversy (by whatever name) sterile and absurd. For example, the anatomy of the vocal apparatus and the history of its evolution are well understood. We are born able to use this apparatus (the baby's first cry) and it is clearly programmed into our genes – its development is innately determined. But no child is born able to speak a language; this is clearly a learned skill – it is acquired. Thus, when considering language and its development, both innate and acquired factors must be taken into consideration.

No reader will be surprised by the assertion above. However, the problem is further complicated by the relationship between our genes and the world. For example, humans and chimpanzees share all but a few genes, and yet the two species are very different in many ways. Why should this be? The answer seems to be that everything depends on the *patterns* among the genes. When a sequence within a gene is changed it will produce a different protein. But what changes the sequence? The answer is a short sequence of DNA called a promoter. Now the important thing about promoters is that although some are directly influenced by other genes, others are influenced by the environment. Thus, nature and nurture really do combine in the development of species, including human beings.

In the present writer's opinion there are, however, two good reasons for continuing to examine the innate–acquired problem in perception. First, individual differences in perceiving are small. What typifies one person's colour vision appears to typify the colour vision of most people. It is therefore interesting to ask, for example, how well a newborn child can discriminate between different hues, or how much experience is required before counting becomes possible. The second point is a very simple one: by setting up experiments to prove whether or not a certain perceptual ability is or is not dependent on learning, a great amount of knowledge concerning perception has been acquired.

After that diversion we may return to Köhler's views. If perceptual experience is a direct reflection of underlying (electrical) brain forces, and if these forces obey physical laws, then it follows that this experience should be as fixed and rigid as the laws of physics demand. Gestalt theorists might have been willing to concede that Köhler's field forces have not been confirmed by subsequent research, but it seems very likely that they would have clung to some form of nativism, given the philosophical origins described at the start of this chapter. Thus, the Gestalt view must be that, while perception can be influenced by attentional processes and the effects of such variables as familiarity, practice, and learning, it is basically fixed in nature. Two tests of this position are possible. First, humans and animals should not be able to reorganize their perceptions. Second, there should be perceptual competence at birth. Both these issues have been addressed in experimental investigations. The obvious theoretical importance of these questions – whether perception is rigid or flexible, and the degree to which it is innate rather than acquired – has led to much research. Animals and humans have been subjected to a number of procedures in which their sensory inputs were distorted or blocked. The severity and bizarreness of these manipulations has varied greatly.

Mild sensory distortion

This has frequently been employed as a test of the flexibility of perception. Such mild distortion has taken the form, for example, of having people wear tinted lenses for long periods. After a time (as any wearer of sunglasses knows) the tinted world reverts to normal: one has adapted to the slight change in the nature of the light entering the eye. Those who have worn prisms that tilt the world in a certain direction also report complete adaptation to this distorted input.

Severe sensory distortion

This can be achieved by lens or mirror systems that completely invert the world. Here reports of adaptation must be treated with extreme caution. It is obviously difficult to know just what complete reorganization of inverted vision would look like, and many reports are extremely ambiguous. Certainly there is *adjustment* to severe distortion: in one famous experiment (Kohler, 1955) the subject was eventually able to ride a bicycle while wearing inverting lenses, but it is not clear whether this was accompanied by phenomenal re-inversion of the world.

Many animals are able to recover from moderate sensory distortion, but in extreme cases, such as when the eyeball of an amphibian was loosened, rotated through 90° and then replaced in the socket, no adaptation took place. Flies with their heads rotated 180° fail to show any adaptation to the consequent inversion of their visual inputs (for an account of some of these dramatic experiments, see Sperry, 1951).

The experimental literature in this area is too large to review comprehensively. At this point we will simply assert that: (a) animals and humans can adapt to many forms of mild sensory distortion; (b) 'higher' animals, such as monkeys and chimpanzees, are better at adapting (are more flexible) than 'lower' animals, such as flies, amphibians, and chickens; (c) it is doubtful whether any species can completely adapt to very severe distortion (although such a statement risks being tautological). Readers wishing to make their own assessments of this literature should consult some of the original publications cited above and in the Endnotes to this chapter.

Our overall assessment of the evidence from distortion studies is that it suggests that basic perceptual organization is relatively inflexible. This accords with the nativist stance taken by Gestalt theorists, although it must be added that some contemporary workers might disagree with this conclusion; many would argue for the old Scottish verdict, 'not proven'.

Animal deprivation experiments

These no longer seem to hold the promise they once did. Rearing an animal without, say, vision, and then testing its visual perception is probably not the way to discover whether visual capacities are innate: the animal is an abnormal animal; suddenly acquiring vision may be frightening; the animal may not be motivated to do well with its new sense. These problems may be insuperable.

Human deprivation studies

Such studies have been used to compare the visual abilities of those who have recovered from blindness with those of the normally sighted. Here, too, problems of adjustment and motivation are to be expected, and this is borne out in many reports. Here is an example of a well-known case:

> The patient S.B., described by Gregory and Wallace (1963), was a man blind from birth. It is important that, although diagnosed as technically blind, S.B.'s eyes remained sensitive to light. Whilst his corneal scarring prevented him from seeing anything of the world, sufficient light reached his retina to keep the cells alive. At the age of 55, S.B. received two corneal grafts, giving him the first chance to see in over 50 years. What was the experience like? Gregory and Wallace, who visited S.B. the day after his operation, report S.B.'s description of the moment his bandages were removed. He explained how he could see something before him, but did not know what it was. When the object spoke, S.B. realized that he was seeing the face of his surgeon.

Gregory and Wallace observed S.B.'s reactions to a number of situations and tests. Three things initially surprised them. First, S.B. was able to tell the time from a wall-mounted clock. This was eventually explained when S.B. told them that he used a watch made especially for the blind and having raised numerals: he was fast and accurate when doing this. It looked very much as though this skill was an example of a cross-modal transfer from touch to vision. Second, S.B. correctly named a magazine he was shown as 'Everybody's'. When questioned, he said that he was able to recognize the first two letters of the magazine's title and guessed at the rest. Once again, cross-modal transfer seems to have been at work, as in his youth S.B. learned to read by touch the embossed letters on brass nameplates. The third surprise was that S.B. performed perfectly on the Ishihara colour vision test. The test comprises a number of coloured plates, each containing a number formed from coloured dots surrounded by other dots designed to confuse the colour vision system. Many years later, the present author and his student (Gordon & Field, 1978) were able to explain this feat (it is a feat, because the present author, having tested more than 100 people on the Ishihara, has found that even those with perfect colour vision sometimes find some of the patterns difficult to detect). Gordon and Field showed that the Ishihara becomes an easier test if the dots are blurred somewhat. If S.B.'s post-operative visual acuity was less than perfect – which was almost certainly the case – then his impressive performance becomes easier to understand.

We have explained what S.B. could do with his regained sight. However, he had many difficulties with other forms of perceiving. His judgement of heights was extremely poor and he felt that he would have been able to step onto the ground from his bedroom window, which was 30–40 feet above the ground. And, perhaps surprisingly, he did not respond in the normal manner to some of the distortion illusions, such as the Müller–Lyer (shown in Figure 2.8) and did not experience the usual perceptual changes induced by drawings of hollow cubes, such as the Necker cube.

Months after his operation, S.B. was constantly afraid of traffic. He had difficulty in recognizing many objects in his environment, and seemed bored during a sightseeing trip to London. It became increasingly obvious that S.B.'s dreams of regaining his sight were not turning out as he had hoped.

One of the most interesting of Gregory and Wallace's observations concerning S.B. happened when they took him to the London Science Museum. He took a long time to recognize a large saw, he could not recognize a model windmill and he was unable to recognize a lathe in a glass case, even though he was deeply interested in tools generally. However, when he was allowed to touch the lathe, he named all the parts correctly, saying, 'Now that I have felt it, I can see it'.

> The end of this case is a sad one. After a time, S.B. increasingly returned to the life of a blind man, getting little pleasure from film or television, and frequently failing to put the lights on at night. He became depressed and at the age of 54 was dead.

The newly sighted are not simply people who have lacked vision, but people who have learned to live with their other senses. In this they are atypical perceivers when their vision is restored. What does seem to be generally true is that such people have great difficulties in organizing and making sense of their new visual world. A nativist would not be dismayed by this discovery, for reasons such as those outlined earlier. However, where the nativist position gains some limited support from these studies is in the general finding that most subjects who have been studied appear to perceive lines, edges, brightnesses, and colours without difficulty (see e.g., von Senden, 1960; Gregory & Wallace, 1963; Valvo, 1971). Equally striking is the fact that in most reports it appears as though the organization of the world into figure and ground takes place quickly and spontaneously, as the Gestalt theorists would have predicted. This evidence suggests that at least some of our visual capacities are innately organized and require little or no learning. Clearly, the above reviews have been biased to studies of humans and the 'higher' primates, for obvious reasons. It is important to say at this point that a very strong version of nativism can be defended if discussion is restricted to the perceptual abilities of simpler organisms. Many creatures show complex behaviour the moment they emerge into the world, and as this may involve things as complicated as flying through the environment, it is clear that their visual powers must be intact from the start: insects mostly do not bump into things (except artificial surfaces such as glass windscreens). There is no doubt that perception can be innately organized, but the interesting question is whether this is the case of humans and similar species. Shortly we shall assert that there are good reasons for believing that a major part of our ability to perceive is innately determined. While this evidence could be marshalled to support *any* nativist theory, it remains true that it was the Gestalt theorists who put forward the most explicit and influential version of nativism and provoked much of the stimulating research that we have attempted to outline.

Infant perception

At the start of this section the author must express his thanks to his colleague and friend, Dr Alan Slater of the School of Psychology at Exeter University. Dr Slater is known internationally for his work on infant vision. He has generously shared his knowledge and judgement when advising on the contents below.

The extent to which infants arrive equipped to perceive the world promises to be, as we have stated earlier, the best way of approaching the old innate–acquired controversy. The Gestalt psychologists were nativists, as we have shown. However, by the 1950s the pendulum was swinging more towards the empiricist (acquired) position. Two important theorists were largely responsible for this shift. Piaget (1954), having studied the perceptual and cognitive development of his own children, concluded that perceptual abilities were extremely poorly developed at birth. Similarly, Hebb (1949) argued that human perceptual development was a long slow process in which experience slowly shaped the ways in which humans perceived the world. In what follows we shall review some research findings that have cast doubt upon the strong empiricist position.

As was discussed in Chapter 1, developments in techniques have often been followed by rapid gains in knowledge. This is nowhere truer than in the field of infant perception. The main problem here is the obvious fact that newborn infants cannot communicate directly with us. How, then, can this problem be overcome without putting the infants at risk or stressing them in any way? A number of new techniques have helped to overcome this problem. Using special cameras, it is possible to shine a light into the infant's eye (often using infrared). This is reflected from the back of the retinal and the image superimposed over the pupil. Thus, the image has passed through four refracting surfaces – the cornea, the lens, and back again. From the shape of this image it is possible to tell the momentary state of the infant's visual accommodation – at which distance the lens was focused. Another technique involves placing electrodes on the infant's skull above the visual cortex to record potentials that will show whether or not the infant is seeing a reversing grating (a display of black and white stripes with the stripes repeatedly changing from black to white, and vice versa) – do the evoked potentials detected via the electrodes show any rhythmic changes in response to stimulation? If so, the gratings can be made finer and finer until the cortical responses stop. At this point, estimates of visual acuity can be made. For more material on the infant's visual acuity, see Campos et al. (2000).

Perhaps surprisingly, the technique that has yielded most data concerning infant visual perception is one of the simplest. An infant is shown a display for a few seconds. Infants are attracted by novelty and it is easy to record whether or not they are looking at the display. After a number of repetitions, the infant appears to grow bored by the display and no longer looks at it. This is known as the *habituation* stage. At this point the display can be changed in some manner. Does the infant start to look at it again? If so, the infant has detected the change: it has *discriminated* between old and new stimuli. This simple technique, including some recent modifications, such as allowing the infant to choose between two simultaneously present stimuli, has solved the problem of communication between infants and researchers. Because such techniques are entirely harmless, it has proved possible to use them with newborns as young as 5 minutes after birth. This is a very important advance. What follows is a selection of findings from this newly invigorated research area.

The newborn's eyeball is smaller than in adults, and the fovea – the site of maximum acuity – is poorly developed. Typical figures from newborns are one to two cycles per degree, about 20/600. By contrast, good adult acuity is about 30 cycles per degree, about 20/20 vision (these figures are taken from Slater, Field, & Hernandez-Reif, 2002). The newborn's vision is about as good as an adult cat's. However, cats get by very well and, as the newborn interacts mostly with the nearby environment, particularly its mother, this is no great handicap.

Newborn infants will choose to look at a 3D object, rather than a photograph of that object (Slater, Rose, & Morison, 1984). They are also sensitive to orientation. When habituated to a black and white striped pattern (a grating) presented obliquely, the infants responded preferentially to a new orientation of the grating (Atkinson et al., 1988).

Size constancy has long been a problem for psychologists and there have been numerous theories concerning its acquisition. It will also be discussed at other places in this book. Briefly, if an object moves away from us, it does not appear to shrink, even though the visual image size is diminishing rapidly (as a function of the square of the distance). Without this constancy it would be impossible to perceive a stable 3D world. Remarkably, it has been shown that newborn infants manifest size constancy. For example, Slater, Mattock, and Brown (1990) showed newborn infants either a small cube or a large cube on preliminary trials, and then the cubes were presented with the large cube farther away and the small cube nearer to the infant, so that their retinal image sizes were identical. The infants looked at the cube they were not familiar with – they preferred its novelty. In other words, the infants had somehow perceived the true object size of the cubes in the preliminary trials. This is a very significant finding. Quite how newborns are able to achieve size constancy will need a lot of thought.

Quinn et al. (1993) found that at the age of 3 months infants group patterns according to the Gestalt laws of proximity (see earlier sections of this chapter for illustrations of these principles).

Figure 2.10 shows a display of the type investigated extensively by Kanizsa (1979). Although no actual contour is present in the figure, most adults normally see it as containing a square – the so-called 'subjective contour'. In a series of experiments, Ghim (1990) found that 3 and 4 month-old infants organize the pattern in the same manner. When a rod is moved backwards and forwards behind a central occluder, so that only the ends of the rod were visible, Kellman and Spelke (1983) found that 4 month-olds later preferred to look at a display of the short ends of the rods – as if they had already perceived and habituated to the whole rod. However, newborn babies do not show this effect: after habituation they prefer to look at the long rod. Slater (op. cit.) concludes that the perception of object unity requires a period of development in infancy.

Newborn infants shown the following moving patterns: (a) a face-like pattern; (b) a partially scrambled face pattern; or (c) a pattern with all the

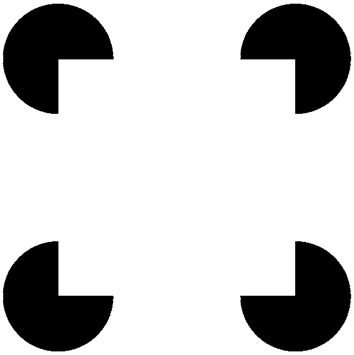


Figure 2.10 A subjective contour.

face elements fully scrambled; followed the patterns in the above rank order (Goren, Sarty, & Wu, 1975; Dziurawiec & Ellis, 1986).

Slater et al. (2000) found that when presented with photographs of faces previously judged by adults as being attractive or not attractive, infants look preferentially at the attractive faces. This is true of infants less than 3 days old, and occurs when the faces are African-American or Caucasian, young or adult, male or female. This is another remarkable discovery.

A long review article by Campos et al. (2000) has the witty title, 'Travel Broadens the Mind'. The focus of the article is the development of locomotion in infants and its impact on psychological functions. Campos and Anderson describe results first reported by Lee and Aronson (1974), who had designed a very ingenious piece of experimental apparatus. This was a room containing three walls and a ceiling. The room was textured and had a soft floor. The special feature of this room was that it could be moved on rails relative to the floor. Infants as young as 13 months, who were just learning to stand, were placed in the room facing the back wall. When the room moved, the infants fell over, showing that they were sensitive to 'optic flow' (the topic of optic flow will be dealt with at greater length in the later chapter on the work of J. J. Gibson). Butterworth and Hicks (1977) later showed that much younger infants, who had not learned to stand, showed similar responses: pressure-sensitive devices in the infants' chairs detected these.

So far in this section, most of the evidence supports the Gestalt psychologists' intuitions concerning innate factors in perception. However, the situation is more complicated than this, as the next paragraphs will reveal.

Any readers of this book who are parents will agree with the assertion put forward by Campos and Anderson in their review, namely that the ability to move marks a very important stage in the infant's development. Before crawling is possible, the infant must be relatively passive, in motor terms. When crawling begins, things change dramatically. For example, instead of beaming at the infant in its cot and making appreciative sounds, it becomes necessary to communicate with it at a distance. The carer (the reason for using this general term will be self-evident) must make loud negative sounds when the crawler approaches danger (the top of a staircase, an electrical outlet, an irritable cat). On their part, crawlers change too. They become more wilful, show distress in the absence of a carer, exhibit more anger and glee and become more liable to initiate simple games with the carer. They have joined what may be called 'the awkward squad'.

Campos et al. (1992) studied the behaviour of infants placed on a 'visual cliff'. This comprises a strong sheet of glass placed over a textured surface. On the 'safe' side of the cliff, the textured surface is just below the glass; on the remaining area, the textured surface is some feet below the glass. An infant can be placed on the safe part of the cliff and then its mother encouraged to call the infant across to her. In their study Campos et al. studied infants who had crawled early, at a normal age, or relatively late in their lives. The findings were clear: it was the amount of previous crawling behaviour that caused infants to hesitate before crossing the deep side of the visual cliff. It is well known that competent adult animals, such as goats, hesitate at the cliff edge of such displays. Clearly, the infant who has crawled is becoming more visually competent.

Campos et al. (1992) found from carers' reports that once an infant started to crawl, he or she became more enthusiastic at initiating interactive games such as 'peek-a-boo' and exhibited more glee when engaged in the games.

Campos et al. (1997) studied the responses of four groups of infants to their mothers' pointing and looking behaviour. The groups were prelocomotor infants, infants with 1–4 weeks of locomotion, infants with 5+ weeks of locomotion and infants who could toddle with the aid of 'walkers'. Toys were placed in a 5 foot square experimental area, on the floor or at or above eye level. The experimenters drew the individual infant's attention to one of the toys by saying, 'look over there' whilst pointing to a toy. The results were very clear-cut: the more locomotor experience the infants had had, the more accurately they were able to follow the experimenter's directions.

As a final example of recent research, we shall describe an ingenious and highly interesting study by Meltzoff (1995). This involved showing 18 month-old infants an unsuccessful effort. An adult attempted to perform an act of reaching for an object, but he stumbled: his hand either under- or overshot the target, or he stumbled several times on the way. Infants who watched these failed attempts later performed the task at a higher level than controls who had not watched the adult. Rather than imitating the adult's clumsy performance, the infants chose to copy the adult's *goals*.

To summarize: there is now a large and growing published literature to show that the Gestalt psychologists were largely correct in assuming that many perceptual functions are present at or shortly after birth. There is also a large literature showing the vital role played by experience in the effective use of perceptual information in *dealing* with the world and learning its meanings. Although the newborn shows some extraordinary perceptual abilities, its world must be very different from our own. Perhaps its experience is like ours when we wake up in a strange room: for a moment, everything can be seen, but nothing makes sense.

We have been able to do only scant justice to the quality and depth of modern thinking concerning infant development. In the author's experience as an outsider, this is a most impressive and fruitful area of modern experimental psychology.

It is impossible in a short chapter to do justice to all the research into Gestalt phenomena that has been published since the pioneering work by the founders of the movement. This section will offer a brief review of a selection of subsequent researches, with a greater emphasis on the most recent work. The review will concentrate on those findings that, we may guess, would have seemed exciting or important to the pioneers of the Gestalt movement.

Subsequent research on some of the Gestalt principles

Grouping by similarity

Beck (1966) subjected one of Wertheimer's laws, grouping by similarity, to an ingenious test. Simple shapes were used to generate textures. A display comprised three sections, each formed from multiples of a simple shape. The observer's task was to indicate which of the two boundaries between the textures was the most natural or salient. Interestingly, observers' choices of boundaries did not strongly agree with their judgements of the relative degrees of similarity between the three sets of elements. Grouping in terms of lightness or colour is easily predictable from relative similarity; grouping in terms of shape is unfortunately not as simple as this: we may conclude that this important rule concerning grouping is still not fully understood.

Symmetry and information

The concept of symmetry is important in Gestalt theory, and is another of Wertheimer's laws. Attneave (1955) brought a new way of thinking to the problem of symmetry, namely the application of information theory. Attneave's idea can be described by referring to the familiar game of 'battleships', in which contestant A draws a shape on graph paper and opponent B attempts to discover the shape by naming squares and asking whether each does or does not contain a fragment of the shape. The game is to find the shape in as few moves as possible. When the shapes used in the Attneave experiment were symmetrical, they were guessed correctly in less time than when they were not. We shall return to this experiment later in the chapter.

Research interest in symmetry has persisted and has led to some very interesting ideas and discoveries. A number of theorists have approached the perception of symmetry from an evolutionary point of view. The fact is that many living things show symmetry, particularly bilateral symmetry. This is as true of humans as it is of lobsters. Further, many objects in the world – fruits, for example – are approximately rotationally symmetrical. As a result, they will give rise to images on the retina that are bilaterally symmetrical. In contrast, non-living things may be highly asymmetrical: think of rocks, lumps of ore, clouds and cascades. A perceptual system tuned to symmetry might have considerable survival value.

Some of the most interesting and provocative recent findings concerning symmetry are those reported by the Danish biologist, Anders Pape Møller. Møller has argued as follows. In a symmetrical organism, perfect symmetry is optimal in the sense that its development represents the ability of the individual to generate the same phenotype under the varying conditions of the environment. In this sense, degree of symmetry represents the genetic fitness of an organism. If this idea is correct, then we might expect that highly symmetrical organisms should be favoured in sexual selection. A selection of Møller's findings will show how research tends to confirm this prediction.

The tail feathers of the swallow (*Hirundo rustica*) show differences in symmetry in different birds. The symmetry of these sexual ornaments was manipulated. Møller then found that females prefer those males having greatest symmetry (Møller, 1992). Other studies of the swallow showed first that the more symmetrical females laid their eggs earlier (an advantage) and were preferred by males; similarly, the more symmetrical male swallows were more successful in acquiring mates (Møller, 1994). The bumblebee (*Bombus terrestris*) prefers symmetrical flowers and these produce more nectar, providing better rewards for the bee. Experimental manipulation of symmetry in flowers also affected the behaviour of the bees in the predicted direction (Møller, 1995).

Although these findings are very interesting, the species studied are very different from humans in structure and complexity. Is it possible to demonstrate any comparable relationships between genetic fitness and preference for symmetry in humans? The answer is yes. Consider the case of the female human breast. For a given age group, breasts clearly vary in size. More interestingly, for our purpose, measurements show that breasts also vary significantly in symmetry. Following the animal studies above, it might be argued that maintaining breast symmetry through variations in the environment provides a 'health certificate'. Is this true? The answer appears to be yes. It has been found that those women in two different cultures, those of Spain and New Mexico, who possess higher degrees of breast symmetry have a greater chance of having children and show greater fecundity overall:

> Males that acquire mates with low levels of breast fluctuating asymmetry will thus tend to sire daughters with little breast fluctuating asymmetry. This will provide choosy males with a sexual selection advantage because their daughters may experience higher mating success, earlier reproduction, and higher fecundity.
>
> (Møller, Soler, & Thornhill, 1995)

It is obviously not easy in modern times to defend the idea that there is any evolutionary selection currently occurring, as Møller would surely agree. The basic selection described above may have taken place in our distant evolutionary past. And yet, how interesting to find symmetry and its possible evolutionary advantage appearing in such a markedly different context from that in which Wertheimer worked 70 years ago.

Work by Gaetano Kanizsa and colleagues

One of the present author's pleasures is to re-read the selection of papers by Gaetano Kanizsa, *Organization in Vision* (Kanizsa, 1979). Kanizsa was a distinguished experimental psychologist at the University of Trieste. He was also an artist. This adds to the appeal of his work, much of which is illustrated with fascinating perceptual demonstrations – maintaining the Gestalt tradition. He was not, however, a rigid adherent to all aspects of the general Gestalt theory, mainly because he saw perception as a constructive process capable of going beyond the information given by stimulation (a theoretical position which will be examined more fully in a later chapter on empiricism). Although Kanizsa believed that seeing and thinking are related, he recognized that they must be distinct processes. Using the familiar Gestalt tactic of the striking demonstration, Kanizsa provides direct evidence to support his position. Note the different impression given by the two cubes represented in Figure 2.11.

The original Gestalt theorists rarely sought to quantify the phenomena that interested them. Here is an example, by one of Kanizsa's colleagues, of how the earlier Gestalt theorists might have quantified something of interest to them. The effect relates to figure and ground separation, which has a prominent place in the Gestalt theory.

Note the two upper patterns in Figure 2.12. In one, the smaller shape is generally seen as lying in front of the larger one; the reverse is true of the other. What causes this change? Petter (1956) discovered the rule governing this phenomenon: it is delightfully simple and deserves the name 'Petter's law'. The fact that we can see such displays as *overlapping* separate figures implies that we have supplied something extra to the displays. Kanizsa would call the conceptually necessary, but perceptually invisible, lines needed to complete the figures, *amodal contours*. Measuring the lengths of these contours is all that it takes to demonstrate the operation of Petter's law. For what is found in examples such as Figure 2.12 is that *the figure*

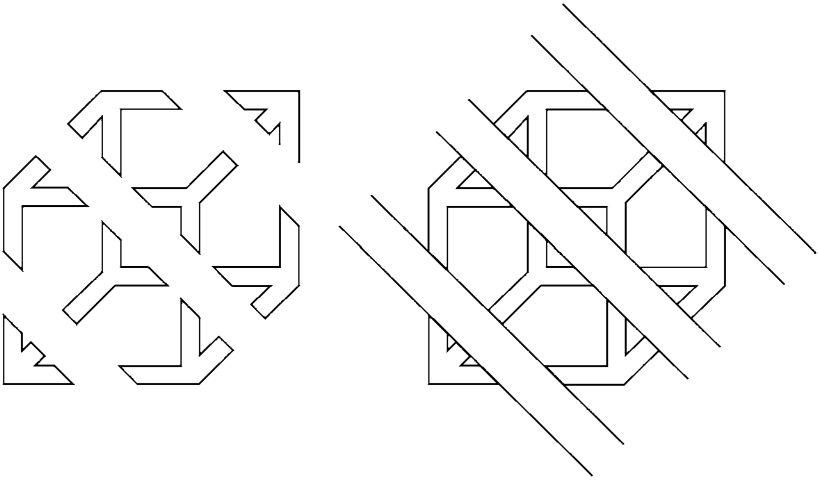


Figure 2.11 Kanizsa's demonstration of the difference between thinking and seeing. The cube on the left can be imagined, but it is hard to see it as such. In the cube on the right, amodal completion takes place and the perception of a cube behind the three stripes is effortless. (From *Organization in Vision*, G. Kanizsa. Copyright © 1979 by Praeger Publishers. Reproduced with permission of Greenwood Publishing Group, Inc., Westport, CT.)

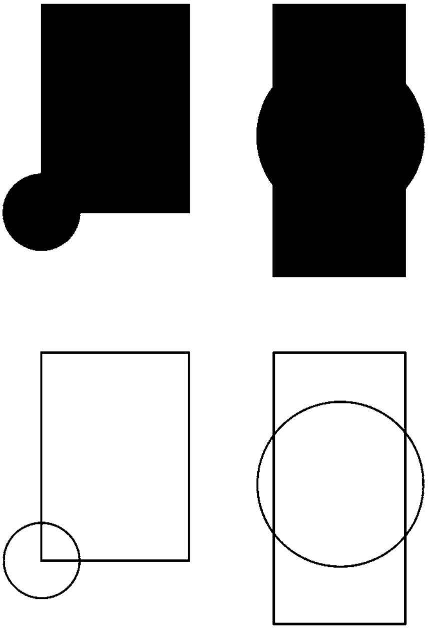


Figure 2.12 Petter's law. It can be seen that the direction of overlap favours the shorter amodal contour length required for completion.

requiring the smaller length of amodal contour for its completion will appear in front.[4] What an elegant finding, and how exciting to discover yet another example of a minimum principle at work in visual perception – assuming that creating an amodal contour requires neural work. With hindsight, we can regret that the early Gestalt workers did not do this simple piece of measurement. It was all that was needed to find a quantitative relationship between figure and ground, and might have given impetus to further quantification.

The minimum principle

Restle (1979) set out to quantify some effects discovered by Johansson (1950, 1964, 1977). Over a number of years Johansson has examined the ways in which subjects see simple moving dot displays. In a typical experiment, three or more dots move smoothly and rhythmically back and forth across a display. The dots may move at different speeds, in various directions, and may be in or out of phase. Johansson found that the perception of these displays is dynamic in the sense that the dots are seen to be connected in some way, or to represent the ends of rigid structures. In one display, for example, two outer dots move up and down in opposite directions while a central dot remains stationary. What one sees in this case looks likes a rigid rod swinging around a central pivot. When a row of dots move different distances, the inner ones moving furthest, one sees something like a skipping rope swinging in and out of the plane of the display. The effects are beautiful and very compelling.

The questions Restle asked were, how can the movements of the dots be described or coded *objectively*, and how do these descriptions correlate with what viewers see? For example, one way of coding the displays would be to treat each dot as independent. Then one parameter will be required for a dot's starting position, one for its amplitude of movement, one for its phase (when it starts to move compared with other dots), one for its angle of tilt relative to the vertical, and so on. In such a manner one can produce an objective description of the movement of the dots, one that could be used, say, to program a computer to drive the displays. But it is also possible to treat the dots as combinations. For example, if a number of dots move together in phase and across the same distance, then a single set of parameters will code the motion of one of them and additional parameters will code the spacings and repetitions. This alternative way of coding will be equally objective, but it will of course be different.

Restle did an exhaustive coding exercise in which the various hierarchies of dot clusters used in the Johansson displays were coded, the dots

---

[4] This law can be overridden by meaning: if the amodal contour would complete say, the tail of a cat, it will tend to be seen in front whatever its length.

treated as independent or linked in some manner. The results of this exercise are highly intriguing. Not surprisingly, the coding of dots as independent events requires the maximum number of movement parameters. When dots are treated as groups, the number of parameters required to code the movement drops, and the more intimate the grouping, the fewer the number of coding parameters required. The exciting finding is this: if one wishes to predict how the moving displays will be seen, then the movement configuration requiring *the fewest objective parameters* is the best bet. And the greater the discrepancy between the number of parameters needed for independent coding and for the coding of a particular grouping, the stronger is the tendency for observers to report the latter.

For now, we may simply note that the tendency to see the displays as simple, grouped, and coherent is exactly what the Gestalt theorists meant by Prägnanz. The difference is that the Gestalt theorists proposed the principle (and gave examples) but offered no means of quantifying it. Here is an example where Prägnanz, operating as a minimum principle, is clearly at work. But now we are offered a possible objective measure of its effects. This is an important development. We shall say more about this approach at the end of this chapter.

Shepard's apparent motion experiment

Here is a brief description of a remarkably interesting experiment by the American psychologist, Roger Shepard. Some of Shepard's best-known researches have been into the nature of internal representations. This has led him to ask certain questions concerning Gestalt principles of organization. Shepard agrees with the idea, described earlier in this chapter, that visual systems attuned to symmetry and other principles of perceptual grouping have survival value. He goes further, however. The world is in three-dimensional Euclidean space. Objects move in this world, but their movement is constrained by an associated kinematic geometry. For example, each member of a pair of gloves lying on a flat surface is a mirror image of the other. No rotation in the plane of the surface can make them congruent, so that their outlines exactly match when one is placed over the other. Either the left or right glove must be rotated in the third dimension for congruence to be achieved.

Early humans would not have survived without the ability to manipulate objects, and in doing so they would have been constrained by the laws of kinematic geometry. Shepard's intriguing idea is that this evolutionary past led to the constraints associated with these laws becoming internalized in our visual systems.

There are an infinite number of ways in which a rigid object at position A can be moved to position B in three-dimensional space. However, there exists one simplest, most economical way of effecting this displacement, and this was published as a theorem by the mathematician Chasles (1830). This theorem proves the existence of:

> . . . a unique axis in space such that the object can be moved from A to B by a rotation about that axis, together with a simultaneous translation along that same axis: a helical twist or 'screw displacement'.
>
> (Shepard, 1984)

At this point the reader should be reminded of some of the material included earlier in this chapter in the review of the law of least action or the minimum principle.

From these theoretical ideas Shepard designed an experiment on apparent motion (see Shepard, 1984), one form of which was described earlier in our discussion of the phi-phenomenon. Pairs of outline polygon shapes were presented in rapid alternation. In each pair, one polygon represented a different view of the other after it had gone through various transformations and combinations of transformations. For example, one member of the pair might represent: a rotation of the other in the frontal plane; a rotation and a change of size, representing a shift in the third dimension; or a mirror reflection. And so on. These different transformations are shown in Figure 2.13.

The question now of course is, what path will observers see the object take during its apparent movement? The answer is:

> . . . in each of these cases, if the rate of alternation is not too great, the motion tends to be experienced as the rigid transformation prescribed by Chasles's theorem . . .
>
> (Shepard, 1984)

Once again, here is a very important and thought-provoking result.

The most recent work on simplicity and the minimum principle

Since the second edition of this book was published, the author's interest has been drawn again to the minimum principle. Two theoretical articles were the cause of this renewed interest: those by Chater (1996) and van der Helm (2000).[5]

What follows is emphatically *not* a précis of these articles or the work they refer to. This is a very complex area of mathematics and the present author is not a mathematician. In fact, the few pages below were written only after many hours' hard reading. Mathematically competent readers may consult the references cited. The Web is another excellent source of information

---

[5] The author expresses his thanks to Dr van der Helm for sending him a pre-print of his important article.

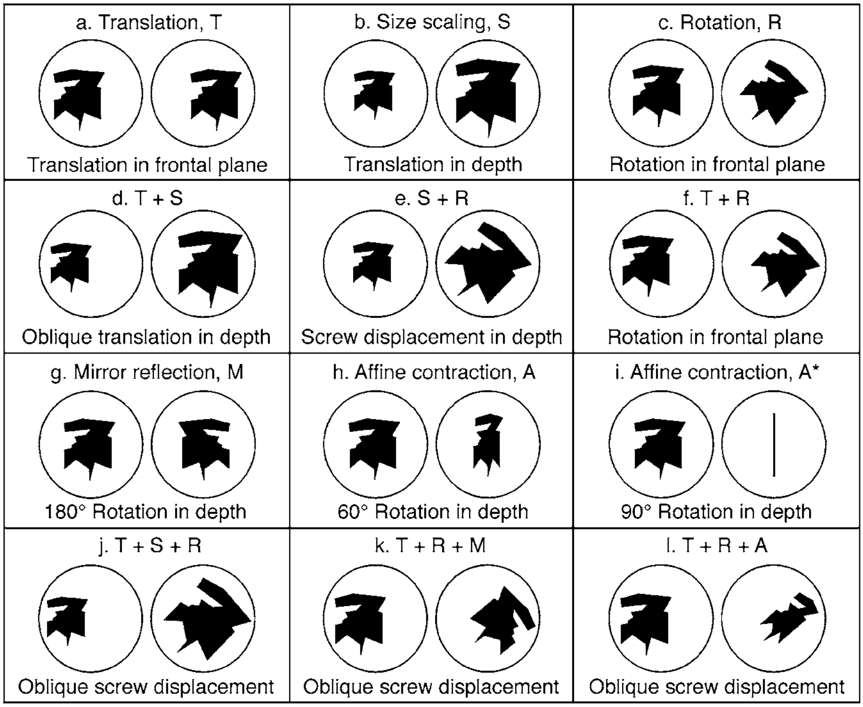


Figure 2.13 Shapes used in apparent motion studies by R.N. Shepard. Each pair of shapes represents a type of transformation. When the shapes are alternated, the apparent motion follows the minimum paths predicted by Chasles's theorem. (From Shepard. Copyright © 1984 by the American Psychological Association. Reprinted with permission.)

concerning recent developments. What follows is an attempt simply to give the flavour of some of this modern work.

In a previous section, we referred to the work of Attneave (1955). Attneave was among the first to attempt to apply concepts from information theory to perceptual problems. This is now the place to say something more about information theory itself.

We all have an intuitive idea about information – it tells us something. But information is available in many different forms: speech, print, television and Morse code, to cite a few examples. What have these in common? How might it be possible actually to *quantify* information in a way that is independent of the medium via which it is transmitted?

In the 1940s in America, mathematicians such as Weiner and Shannon were working on problems of encryption and decoding. Then, within this context, Claude Shannon published a landmark paper entitled, 'A Mathematical Theory of Communication' (Shannon, 1948). In essence, Shannon's insight was as follows. Suppose one has a coin with heads on both sides. The coin is now tossed and lands . . . heads. Has any information been transmitted to a viewer of this incident? The answer, of course, is no. Now imagine a fair coin being tossed. A gambler knows that this might result in a head or a tail. He or she is in a state of *uncertainty*. The outcome resolves this uncertainty. A similar situation exists when a die is thrown. When it comes to a stop, a gambler has his or her uncertainty reduced; intuitively, this reduction is greater than that in the situation where a single coin is tossed. Finally, consider the throw of a ball into a spinning roulette wheel. The chances that the ball will end up in any position are one in twenty. When the wheel eventually comes to rest, the outcome reduces the gambler's uncertainty – which was even greater than when watching the coin or the die. Therefore Shannon was led to link information with uncertainty; when an event occurs, the greater its power to surprise us, the greater the information it transmits. This is also known as the 'surprisal' of the event. The question now is, in what units shall information be measured? To gain an intuitive grasp of Shannon's solution, consider the problem of giving a unique code to each of the following eight letters:

A B C D E F G H

A convenient systematic way of doing this is to divide the letters into two groups, the left and right halves. Give each of the left-hand group the code number 0 and each of the right-hand group a 1 (remember that this work was carried out when computers were becoming available, and they work in binary code). Now repeat the process within each subgroup, and again for each of the remaining pairs of letters. The code for each letter will therefore be as follows:

A: 000
B: 001
C: 010
D: 011
E: 100
F: 101
G: 110
H: 111

Note that in order to code eight items we need a code length of three binary digits or bits. The significance of this is that $3 = \log_2 8$ ($2 \times 2 \times 2$, or 2^3). Had we coded 16 items in the above manner the code length would have been 4, as $\log_2 16 = 4$ or 2^4 (note also that if each of the original eight letters had been given a code of four binary symbols, this would have introduced an element of *redundancy* into the coding).

We can now quantify the optimum number of bits required to encode a message of length n when the ensemble of symbols (or events) is equiprobable, as in the examples above. The answer is $\log_2 n$.

In many situations, however, the possible signals or events will not be equiprobable. In written English, for example, the letter 'e' occurs 130 times per 1000 letters on average. By contrast, the frequency of 'j' is only 2 per 1000 letters. (The now-defunct Morse code reflected these facts: the more common letters are coded into fewer signals, so that 'e' is coded as a single dot, while 'j' is coded as a dot and three dashes.)

To deal with unequal probabilities, Shannon's formula becomes:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

For those unfamiliar with the symbols, Σ tells us to sum the probabilities that follow from $i = 1$ (the first of the sequence) to $i = n$ (the last in the sequence). Once again note that the minus sign before Σ deals with the negative values arising when logs are taken of fractions. The formula yields the average amount of information per symbol (H) in a signal in terms of binary digits. In other words, if the original signal is converted into a string of 0s and 1s, this is the average number of binary digits required per symbol or event. Rarer events or signals will require more binary digits than common ones. Readers new to this way of thinking may be helped by the term 'average surprisal', which can be substituted for H in the above formula. This term is in fact used by some theorists.

All communications systems contain *noise*; that is to say, they show activity in the absence of any signals. For example, a radio not tuned in to a broadcast frequency 'hisses' – this is the noise that is an inevitable consequence of the electronic circuitry. In like manner, when a television set is switched on in the absence of incoming signals, the set is not blank; rather, it is covered in small spots of light. This again is the electronic noise in the circuit. One of Shannon's great achievements was to show how signals could be designed to penetrate noise by the use of redundancy, a proof that transformed the science of communications. In order to show the magnitude of Shannon's achievement, we need only record that, before his publication, the largest communications channel carried a maximum 1800 simultaneous voice communications. Twenty-five years later that figure had risen to 230,000 simultaneous voice communications. At the time of writing, a single optic fibre, the thickness of a human hair, is carrying over 6 million simultaneous conversations. Shannon's work was truly revolutionary.

We can now better understand the significance of Attneave's experiment, mentioned in an earlier section. Attneave asked his participants to play a game akin to battleships. When he introduced a symmetrical shape, such as a drawing of an ink bottle, his observers were able to guess the right-hand half of the display faster after having guessed the left-hand half correctly. Thus, the Gestalt notion of symmetry might now be interpreted as reflecting redundancy of information. This is an interesting idea.

To summarize so far: mathematicians developed a method of measuring information in an objective manner. The method relates information to uncertainty. The quantity of information in a sequence or ensemble is measured in binary digits or bits. As long as uncertainties can be measured, the medium via which a message is transmitted is unimportant. It is not hard to see why workers in disciplines such as psychology and physiology found the development of information theory so exciting. Of course, when studying perception, the probabilities of events in the real world will commonly be unknown; that, however, is an empirical problem and one that could be subject to research.

Discoveries in a new area of mathematics have united information theory and computer science. *Algorithmic information theory* was first outlined independently by Solomonoff (1964), Kolmogorov and Uspenskii (1987), and Chaitin (1966). There are many labels in this field of mathematics; we shall use the common one, ‘Kolmogorov complexity’. In other words, we shall concentrate on the work by Kolmogorov and his work on the minimum description length principle. What Kolmogorov complexity offers is a new, powerful, objective measure of complexity.

We shall now attempt to say something about how this work can be used. The approach to be followed will be an indirect one. Consider the following three sample sequences, and imagine that the true length of each is 500 symbols:

(a) 121212121212 . . .
(b) 314159265262 . . .
(c) 018348623853 . . .

Suppose now that we wanted to program a computer to generate each sequence. In the case of (a) a simple program would be somewhat as follows: for i = 1 to 500 do {print (1), print (2)}.

The case of (b) seems much trickier – there appears to be no pattern here. However, alert readers will have spotted that in fact the sequence represents the first digits of the constant, π(pi). There are several formulae available for generating π. One that is used by modern mathematicians comprises approximately 19 terms and has been used to generate π to over eight billion digits. In other words, a program to generate (b) would be longer than that used to generate (a), but not dramatically so.

When considering the sequence (c) it might be hoped that a similar formula to that for generating (b) is available. In fact this is not the case; the digits are from a table of random numbers and the only program that could be used to generate (c) would need to have the form, for i = 1 to 500 do {print (0), print (1), print (8), print (3) . . .}. That is to say, this

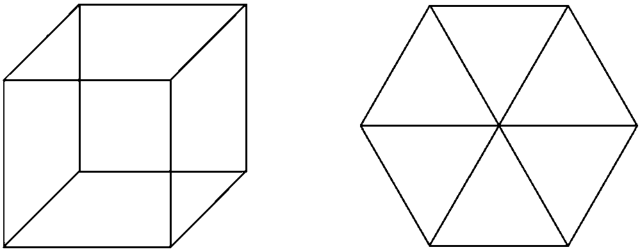


Figure 2.14 Two projections of a hollow cube. One is symmetrical in the plane and remains stable. The other, although equally accurate as a representation of a cube, is not as symmetrical in 2D and tends to be seen as a 3D representation, in which it is more symmetrical.

program would be slightly greater in length than the sequence it generates.

To reinforce the above, consider Figure 2.14. The Gestalt psychologists noticed that of the two patterns in Figure 2.14, the one on the left tends to oscillate between two views of a cube. The pattern on the right may occasionally oscillate, but is generally more stable and unchanging. Their explanation of this in Gestalt terms was that the stable figure on the right is symmetrical and has high Prägnanz. The hollow cube figure on the left is lower in Prägnanz when seen as a 2D pattern. But when it is seen as a regular 3D cube it becomes simpler and more regular – it has higher Prägnanz. This was the Gestalt explanation of why the figure tends to be seen as a 3D cube. In fact, *both* figures are accurate representations of regular cubes, although one represents an unusual view. Following the theme of this section, we can deduce that a program to draw the left-hand pattern in Figure 2.14 would be much shorter that one that could draw the right-hand pattern. (Of course, modern graphics programs commonly have features allowing the rotation of figures in three dimensions, in which case the right-hand cube would require a drawn square to be moved into five other positions; however, it would still require a longer program.)

As a final example, consider a program designed to draw a schematic face, similar to that shown in Figure 2.6. The program would be relatively short. Now consider a program to draw a life-like face; this would be very long.

The length of programs or descriptions is a major concern within algorithmic information theory. In fact it leads directly to a definition of Kolmogorov complexity, which is defined as the size of the shortest program (or algorithm) in bits that, without any additional data, computes a string and terminates (the term ‘string’ is the way in which mathematicians think about the programming problems outlined above).

To date, the emphasis in the area of algorithmic information theory has been on the fundamental aspects of mathematics. Chaitin, for example, writes about what he terms ‘metamathematics’. To date, then, the true home of algorithmic information theory is within mathematics (and is concerned in part with the basic nature of mathematics) and computer science.

Let us, as non-mathematicians, leave these esoteric concerns, for the fact is that of late, researchers in the field of visual perception are beginning to use the theory. A number of workers, including Chater and van der Helm, to whom we expressed indebtedness at the beginning of this section, are using algorithmic information theory, and in particular Kolmogorov complexity, in attempts to solve problems in vision; they are trying to use objective measures in their work. To readers who have stayed with us thus far, we strongly recommend turning to van der Helm’s highly informative paper (van der Helm, 2000). There, readers will find a set of fascinating attempts to consider perceptual problems from the viewpoint of algorithmic information theory, and the article is richly illustrated. It seems fair to say that, to date, there have been no dramatic breakthroughs. But it is now possible to feel real excitement over the fact that the claims made by Gestalt theorists concerning perception and simplicity may soon receive critical tests. In other words, if we can now measure the relative simplicity of rival perceptual outcomes or interpretations, then we can re-examine the concept of Prägnanz. If the Gestalt psychologists are proved to have been generally correct in asserting that perceptual outcomes tend to be the simplest, then this will be a lasting tribute to the power of their intuitions.

Endnotes

- We have occasionally expressed regret in this chapter over the fact that Gestalt psychologists did not pursue certain research problems more vigorously. One of the reasons why is a tragic one. Several key figures in the movement suffered under the Nazi regime. Some were forced to emigrate, to the detriment of their careers. Otto Selz, who worked on thinking, died in a concentration camp. Karl Duncker, famous for his ingenious studies of problem-solving, committed suicide at the outbreak of the Second World War. Mandler and Mandler (1969) give an account of this tragic history.

- Gestalt writings on perception are clear, accessible and very interesting. Readers wishing to learn more about this approach should start by reading original Gestalt documents, particularly the Köhler and Koffka references given in the text. (It may seem strange that a number of the references given below appeared so long after the start of the Gestalt theory. Many of these are English translations that appeared after the Gestalt theory became more widely known and the Gestalt psychologists had moved to the USA.)

- Köhler's field theory of neural action is described in his book, *Dynamics in Psychology* (1940).

- The source book by Ellis (1938) is invaluable and contains some of the best-known replies of Gestalt theorists to their critics. See particularly Koffka's (1915) Reply to Benussi and Köhler's (1925) Reply to Müller.

- Petermann (1932) provides a useful description of Gestalt work on dynamic aspects of perception. Part 1 of this book contains a description of Köhler's views on physical Gestalten and his brain model.

- Bruce, Green, and Georgeson (1996, Chapter 6) contains some beautifully illustrated illustrations of perceptual organization.

- A central theme in this chapter has been that of perceptual organization. There is, in fact, a book with this title written by Michael Kubovy and James T. Pomerantz (1981). The work described therein (and in other publications by Kubovy) should now be read by interested readers. This important work cannot be described briefly; it was therefore decided to omit it from the present chapter, rather than to offer an inadequate account. It is, however, of the greatest interest and relevance.

- A Gestalt psychologist who has not been referred to in the main text, but whose work deserves mention, is David Katz. See, for example, his book, *Gestalt Psychology* (1951).

- Readers who enjoy the style of the best Gestalt writings will also enjoy reading Michotte's (1946) work on the perception of causality. This shares some of the characteristics of Gestalt research and the demonstrations, which are easy to set up, are extremely compelling.

- Further accounts of the work by Slater and his colleagues can be found in Slater and Morison (1985) and Slater (2001, 2002).

- The distinguished psychologist Julian Hochberg attempted to extend and improve some Gestalt ideas on visual perception. See Hochberg (1968, 1973) for an account of some highly interesting perceptual researches.

- The Web is a valuable source of information related to Gestalt psychology. For example, the Society for Gestalt Theory and Its Applications (GTA) is easily located using a search engine such as Google. The society has arranged for a long list of classic texts by Gestalt theorists to be available for purchase.

- Some interesting applications of Gestalt theory to the arts are described by Arnheim (1949, 1956, 1969). Arnheim has also written interestingly on the precise meaning of the term *Prägnanz* (see Arnheim, 1987).

- Readers other than computer scientists will have found the section on algorithmic information theory the most difficult in this chapter. In fact, those who consult the references given in this section will find that it is incredibly more difficult than our description. Much of Chaitin's work, for example, uses the programming language LISP, with which the present author is not familiar. One of Chaitin's equations is 200 *pages* long and has 20,000 variables. This is stern stuff. For further reading, see Chaitin (1990, 1999).

- In the week when this chapter's section on recovery from blindness was being re-edited, *The Times* (25 August 2003) published an account of the recovery from blindness in an American, Mike May. The full account of this work has not yet appeared, but will be published in *Nature Neuroscience*. Mr May lost one eye and all sight in the other because of an accident when he was aged 3. Since then he has lived a vigorous and determined life – he is a former champion blind skier – and has his own Website (www.Senderogroup.com/perception.htm), where he reports on his experiences. Surgeons in California treated Mr May's retina with stem cells to repair the damage. They then replaced his cornea with a graft. He can now see. However, Mr May is having severe problems in dealing with the visual world. For example, although he can deal with simple, basic forms, he has great difficulty with 3D perception. He has great difficulty in deciding whether a particular face belongs to a man or a woman, or looks friendly or hostile. He makes mistakes when identifying common objects. When trying to identify a face, he is commonly forced to base his judgement on particular local features, such as hair length or the shape of the eyebrows. From what we know of the S.B. case, none of this should surprise us. What is important and exciting, however, is the facts revealed when Mr May attempts to perform visual tasks during fMRI scans – a technique clearly not available to earlier investigators such as Gregory, Wallace, and Valvo.

  There are over 15 regions of the brain known to be involved in visual processing. In Mr May's case, only some of these are functioning normally. As *The Times* account suggests, this may be because the remaining regions have been 're-wired' to perform other tasks. As a speculation, we may offer the following detail. There is a condition, associated with forms of brain damage, known as *prosopagnosia*. Those suffering from this condition have difficulty in recognizing faces, even their own on occasion. It is commonly found that prosopagnosia arises after damage to the parietal lobe of the cortex. In the case of S.B. and Mr May, might it be the case that this is one of the regions that was ‘re-wired’ during their prolonged periods of blindness?

- We end these notes with a fanciful speculation. When watching a classical ballerina, we witness an unusual grace in her movements. Remembering our earlier description of the Chasles theorem, could there be a link?

3 Brunswik's probabilistic functionalism

The second theory to be described is probabilistic functionalism, which was essentially the work of one man, Egon Brunswik (1903–1955). Brunswik left no school of followers, his work is rarely cited in modern writings, and no subsequent group of workers has ever assumed the label, 'probabilistic functionalist'.[1] There are, however, good reasons for including this short chapter on Brunswik's work. We shall attempt to justify this claim by describing Brunswik's theory as sympathetically as possible, before offering a number of criticisms.

The remainder of this chapter will cover the following topics:

- Egon Brunswik.
- A general outline of Brunswik's work.
- The terminology of probabilistic functionalism.
- Brunswik's lens model.
- Brunswik's empirical researches.
- An evaluation of probabilistic functionalism.
- Final remarks on Brunswik's theory.
- A more recent development: the empirical theory of vision.

Egon Brunswik

Egon Brunswik was born in Budapest in 1903. As a child he spoke Hungarian and German. His university education took place in Vienna, where he worked under Karl Büller. It was there that Brunswik came to know the Vienna school of logical positivists, whose views were to influence his later career as experimenter and theoretician. In 1937 he moved to the University of California at Berkeley. He died in California at the age of 52.

---

[1] There are, however, Egon Brunswik societies, which can be found on the Web.

Brunswik's work on perception is described in a number of papers in English and German. Detailed statements of his theoretical position are to be found in The Conceptual Framework of Psychology (1952) and *Perception and the Representative Design of Psychological Experiments* (1956), which was published posthumously.

A general outline of Brunswik's work

Much of Brunswik's thinking concentrated upon the relationship between distal and proximal events and what the brain must do when the correlation between these is less than perfect.

It may be useful to begin with a general outline of Brunswik's thinking and its historical context, before considering the details of his work.

Brunswik and the inference revolution

What Gigerenzer and Murray (1987) have called 'the inference revolution' was nearing completion by the time Brunswik arrived in the USA for the final stage of his career. The inference revolution had two effects. First, it became increasingly natural to interpret a number of psychological processes in terms of statistical decision making, in a manner analogous to that involved in testing the truth or falsity of scientific hypotheses. Second, by the 1940s the work of Fisher, Neyman, and Pearson was beginning to provide psychologists with powerful tools for analysing data: the various tests of significance, such as the t-test and analysis of variance, are obvious examples. Between the mid-1930s (when Brunswik moved to America) and 1940, the total number of articles published in the psychological literature reporting tests of statistical significance was a mere 17. By 1960, 5 years after Brunswik's death, such reporting had become the norm. We shall show later how the lack of statistical sophistication among psychologists (including Brunswik) reduced the impact of Brunswik's novel and interesting ideas and made it difficult to refine and extend them.

The statistical nature of cues

One of the most important things that Brunswik did was to extend the notion of uncertainty within the individual perceiver by ascribing it also to the physical world. Brunswik maintained that it is not just that among sensory processes the need for statistical evaluation and decision arises, but that the world itself is an uncertain place. He showed how the cues arising from objects and events in the world are commonly less than perfectly reliable. However, these imperfect cues seldom arise in isolation. Objects and events are frequently complex, and because of this they generate not one but many cues. The problem for the perceiver is how to arrive at rapid and generally valid perceptions of the world based on uncertain information conveyed by these varied cues.[2]

Brunswik's functionalism

Brunswik's functionalist approach was reinforced by the fact that perceivers do, in fact, usually get things right – they would not survive if they did not. And they often get things right very quickly, which has obvious survival value. The question is, of course, how does all this happen? Brunswik's answer was that in order to understand perception we must begin by studying the environment or ecology from which perceptual processes evolved and in which they continue to function. And we must study this ecology in its full complexity, avoiding the artificiality of the controlled, single-variable laboratory experiment. Only if we observe perception under complex, life-like conditions will we discover how it functions under these conditions. His programme therefore called for 'representativeness' to replace control and artificiality, and was designed to allow observers to show the flexibility and subtlety that they needed in their dealings with the everyday world. Whether or not Brunswik's aims were achieved will be the focus of this chapter.

The terminology of probabilistic functionalism

An economical way to introduce Brunswik's approach is to explain some of the most important concepts and terminology appearing in his writings. Brunswik's writings can be difficult to follow on first reading, as will be shown. One source of difficulty is his occasionally difficult terminology: once this has been mastered, the writing becomes much easier to follow. A picture of the general approach will emerge via a description of the particulars.

Distal and proximal cues and the achievement of stability

As we study a perceiving organism, it becomes increasingly obvious that its behaviour is directed not to the pattern of stimulation on the sense receptor, but to the world beyond. Although it is possible to list *proximal* variables or *cues*, such as the sizes and shapes of retinal images, behaviour is directed not to these but to the actual properties of things and events out in the world, to *distal* variables. The researcher's main task is to discover the basis of this achievement. Further, all that we know about our own perceiving (and what

---

[2] Hammond (1966) points out that, as a schoolboy, Brunswik studied the history of the Austro-Hungarian empire in both German and Hungarian and noticed discrepancies between the accounts, and that he may have remembered this later when thinking about the probabilistic nature of knowledge, cues, and the environment.

we can infer about perceiving in other species) tells us that the central achievement of perception is *stability*. For example, although images on the retina are constantly shifting because of head and eye movements, the phenomenal world is a stable one.

The probabilistic nature of cues

The environment, to use Brunswik's term, 'scatters its effects'. The cues that arise in the external world are only *probabilistic* and not fully dependable. But if this is true, what possible basis can there be for the achievement of perceptual stability? An example (not Brunswik's) will reinforce this important point. Suppose we are searching for edible fruit. Let us assume that edible fruit is in fact generally (a) darker, (b) redder, (c) softer, and (d) sweeter. Obviously, darker and redder are visual cues, softer is tactile, and sweeter is gustatory: the environment is scattering its effects. And these cues, the only ones available, are all imperfect – all carry some risk. Not all ripe fruit is red, neither is all red fruit edible. Sweetness often indicates edibility, but some poisonous fruits are sweet. Some fruit is more edible when soft; some soft fruit will be rotten.

Vicarious functioning and the perceiver as an intuitive statistician

What can the perceiver do, faced with such uncertainty? We must remember that for millions of years such problems were, literally, a matter of life or death. Brunswik's answer was that in order to survive, the perceiver must be able to act as an *intuitive statistician*. It is necessary to weigh and combine cues and shift from ones that are not available to others that are. As cues are varied and commonly have reliabilities of less than 1.0, the environment is described as being *vicariously mediated*. The response to vicarious mediation is *vicarious functioning* in the perceiver. Thus, flexibility in perception must be accompanied by flexibility of response: organisms are clearly goal-directed. For example, an experimental animal, prevented from gaining access to reward in a usual manner, will find another solution to the problem. A rat that has learned to run through a maze will swim if it is flooded.

Perception, then, is *uncertainty-geared*. It aims for '. . . smallness of error at the expense of the highest frequency of precision' (Brunswik, 1956). And as perception involves the evaluation of evidence from different sources, the estimation of relative probabilities and decisions about the attainment of goals, it clearly shares many of the properties of thinking. There are differences though. Thinking aims for definite answers; it is 'certainty-geared'. Thinking is deterministic and discontinuous; it is characterized by 'sudden attainment', often following lengthy pauses. These qualities are sufficiently different from perception for the latter to require a special term: *ratiomorphic*. Thus, a clever person will not necessarily be a better perceiver. And illusions commonly persist even when they have been 'explained' to us (Brunswik once referred to this as 'the stupidity of the senses').

The validity of cues

Validity is an important concept in Brunswik's theory. The nature of the physical world, and the structure of the various sense organs, creates relationships between distal and proximal variables. For example, one impressive aspect of vision is its stability. Researchers can measure the size, distance and position of any object in a field of view. From this list of distal variables we can calculate the sizes and positions of retinal images (or, more conveniently, we can take photographs from the observer's position and use these as substitutes for the retina). These new values tell us about proximal variables. In Brunswik's opinion, the task remaining for the researcher is to discover the relationship between the distal and proximal variables, for here must lie the key to the achievement of perceptual stability. The relationship will seldom be perfect and simple, of course: distant objects usually form smaller images than near objects, but very small objects form small images even when they are near.

How might the relationship between distal and proximal variables be quantified? Brunswik suggests that the *correlation coefficient* is the most appropriate measure. The magnitude of the coefficient offers a useful index of the *ecological validity* of a particular cue – retinal image size in the example above. Ecological validity will seldom be perfect, as we have shown, but obviously some cues will be better than others and this will be reflected in higher correlations.

The ecological validity of a cue indicates its potential usefulness for an organism, but does not reveal whether or not the cue is actually used. The researcher must now ascertain whether or not a potential cue has *functional validity*. Consider, for example, the role of the two eyes in stereoscopic vision. It is known that this form of depth perception is based upon the small differences between left- and right-eye views that exist because of the lateral separation between the eyes. The resulting retinal disparity is a powerful source of information concerning an object's position in the third dimension: it has high ecological validity. But it is of no help to a small minority of people who lack the ability to fuse information from the two eyes. For these 'stereo-blind' individuals, the ecologically valid cue has no functional validity. In contrast, people have been shown to base their judgements of the intelligence of others on aspects of their appearance. For example, wearers of spectacles tend to be judged as cleverer, but as this is not in fact true, we can say that spectacle wearing may have high functional validity (it is used as a cue), but low ecological validity (don't trust guesses about intelligence based on appearance).

To summarize so far. Brunswik's writings suggest an analogy between a perceiver and a boxer who is fighting to survive. On no account must the boxer take a hard punch to a critical area, but he (and increasingly these days, she) must always be ready to seize opportunities to attack. It is vital to anticipate the opponent's moves: which is the real threat, which the feint? The boxer needs clues, ways of predicting what will happen in the next fraction of a second. The opponent will inevitably give some hints – movements of the arms and legs, changes of expression, shifts of gaze – but none of these is entirely to be trusted. A well-matched fight is in part a gamble: the first fighter to predict accurately will survive; errors will be punished. And there is no set of rules, no textbook of boxing, that can guarantee success. It all depends on getting things right at the time, with speed rather than precision.

Brunswik's lens model

Brunswik considered this to be an important part of his theory. The model was meant to illustrate how perception involves a kind of focusing: the scattered and mutually substitutable cues arising from the environment must somehow be gathered together for possible use. Perception involves a focusing of cues, it 'achieves' distal objects and it is towards these that responses are directed.

In its original form, the lens model treated perception as analogous to a single biconcave lens (Figure 3.1). Later versions reflect Brunswik's increasing recognition of the importance of central, ratiomorphic processes: habits, evaluations, and predispositions can all influence behaviour. Also, it is (trivially) true that central factors must underlie mutually substitutable responses: we can respond to a stimulus by speaking or pushing a button if

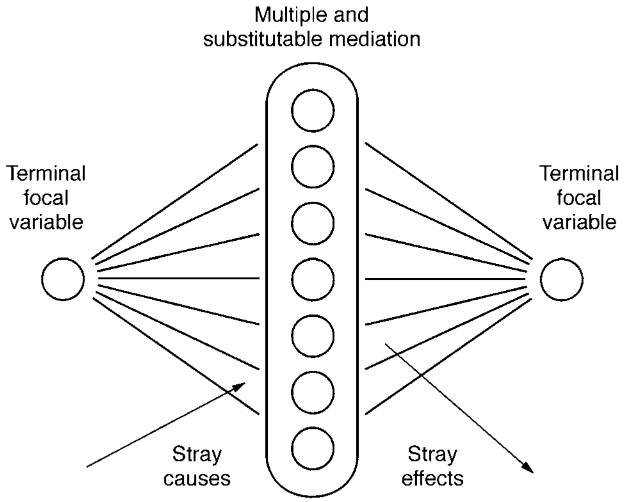


Figure 3.1 Brunswik's lens model.

asked to do so. For these reasons, a pair of biconvex lenses might be more appropriate as a model of perceptual processes.

Brunswik believed that his lens model could guide the quantitative assessment of particular perceptual achievements and thus assist researchers. In an illustrative exercise (Brunswik, 1956, Figure 10) data from a study of size constancy are summarized in a lens model. Photographs of a real scene were taken and the actual and proximal (or photographic) sizes of objects compared. Other cues to distance were recorded. Observers then judged the sizes of objects from the photographs. Correlations between object and image sizes yielded values for ecological validities, and correlations between cues and judgements yielded functional validities. The various cues available in the photographs (e.g., size and vertical position) could then be ranked in terms of their relative importance. Although Brunswik did not do this (he did elsewhere), it was now possible to use the correlation coefficients to trace 'principal rays' through the lens, revealing the basis of the form of perceptual stability represented by size constancy. In other words, people can judge sizes from photographs – the overall reduction in the sizes of depicted objects does not greatly hinder them – and the lens model can reveal the basis of this attainment.

The lens model will not be discussed further. As a way of conceptualizing various aspects of perception, it can be useful. As a means of gaining deep insights into perception, the lens model appears to have little to offer, a conclusion that is reinforced by experience with students who appear to gain little understanding even of Brunswik's own work by concentrating on this single part. It should be said of the lens model, however, that it is one of the earliest examples of this type of thinking in the history of perceptual research. And Brunswik himself thought it important.

Brunswik's empirical researches

Brunswik's style of thinking can become more understandable when his own empirical investigations are studied. Some of these are valuable less for what they achieved than for what they attempted. This review will omit Brunswik's earlier, more orthodox researches and will concentrate on three of his most original publications.

Brunswik's experiments are unusual. They cannot be described as systematic explorations of a group of phenomena, neither do they constitute critical tests of a theory. Rather, they are demonstrations that allow some of Brunswik's ideas to be presented in ways that make them more understandable and memorable. We shall describe only three examples of Brunswik's researches. These will give a fair impression of his originality. The reader should be reminded that, at the time they were published, there was almost nothing like these experiments in the research literature.

Judging coins

In many currencies, coins vary in both size and value; commonly these two variables co-vary – larger coins are worth more.

Brunswik assembled some roughly circular clusters of Turkish coins (he was working in Ankara at the time). The standard cluster comprised 40 two-and-a-half cent coins with a total area of 16 units and a total value of four Turkish units. Other clusters differed in number, area and total value. Observers who were familiar with the coins were then asked to judge a series of comparison displays and compare them with the standard in terms of (a) area, (b) numerosity, or (c) value. What emerged was that judgements in terms of one variable were affected by others (the results are described in Brunswik, 1956), for example, perceived area increased with increased value. This was also true of perceived numerosity.

Similar results were found in another experiment, in which observers adjusted the height of an elongated rectangle in order to match the area of a standard. Here, observers selected heights that were too great: the long thin shapes were perceived as having smaller areas.

These results supported Brunswik's claim that in perceiving the complex world, from which numerous proximal cues arise simultaneously, what is perceived is a *perceptual compromise*. This idea was later refined by Helson (1947), who showed that, for example, the perception of coloured surfaces can be biased systematically by controlled changes in context and illumination. Most recently, perceptual compromise has been invoked by Day in his attempts to explain certain geometric illusions (see, e.g., Day, 1989). Thus, the distortions occurring in the perception of the Müller–Lyer illusion may represent perceptual compromises between the perceived length of the shafts and the perceived extents of the entire configurations, which vary according to the direction of the arrow lines – inwards or outwards.

Perceptual compromises are not the same thing as the powerful stimulus interactions demonstrated by the Gestalt psychologists. They do tell us one thing, however: if the perception of certain stimulus dimensions can be affected by the presence of others, this must be taken into account in experimental research. This is clearly a complicating factor of great importance. If Brunswik's views are generally correct, the study of perception will become increasingly complex and difficult.

Size constancy under real-life conditions

The traditional psychophysical method of measuring size constancy requires the observer to adjust a near stimulus until it matches the apparent size of a distant one. Typically, the measurement takes place in a large room or a long corridor. The variables manipulated include the attitude of the observer, the distance of the far stimulus, the use of one or two eyes, and so on. This of course is the type of design to which Brunswik objected because of its artificiality, its failure to sample the real environment.

In an experiment published in 1944 and reworked in his 1956 book, Brunswik describes a very different approach to the phenomenon. A student was followed outdoors for a period by a psychologist who asked her to estimate the size of that object which was currently dominant in her field of view. From the resulting sample of 174 estimates a number were selected for further analysis, which involved measuring the actual sizes and distances of the judged objects.

The size range of the objects was very large, 105:1, much greater than any that had ever been used in a laboratory study (Brunswik was clearly correct about the unrepresentative nature of much of the research in this area). Brunswik's reporting of his main results leaves much to be desired. The main finding, however, is of some interest: the correlation between object size and image size (expressed in this case as the angle subtended at the eye by an object) was only 0.7 over all objects, dropping to 0.1 when small objects were excluded from the analysis. Thus the ecological validity of image size is low. But the overall correlation between object size and judged size was extraordinarily high: 0.99. Clearly the observer had achieved true and valid distal focusing despite the relatively poor utility of one well-known cue to size constancy.

At this point we await a lengthy discussion by Brunswik on how the sizes of objects in the real world are 'attained': his work has uncovered an important paradox and we await his speculations on this with interest. Disappointingly, we are offered only a brief description of distance cues in general and reference to a proposed experiment using photographs. And the matter is even more serious than this. When, later, Brunswik and his original observer sat in a room, the observer was able to make the same judgements from memory and produced essentially similar results. This finding must be considered to throw some doubt on Brunswik's general views on perception.

This is not the place for a lengthy discussion of the perception of size: interested readers should consult the admirable review in Kaufman (1974, Chapter 9). But there is a problem here that Brunswik does not face up to. A common technique used in the study of size constancy is to provide the observer with a variable stimulus within arm's reach. The observer looks at a distant target and adjusts the variable so that it matches it. Various instructions are used: the observer may be instructed to try to achieve a 'retinal match', that is, to make the two stimuli subtend the same visual angle, or he or she may be asked to match the actual size of the distant object. Interestingly, it is hard to achieve true retinal matches because of the tendency to respond in terms of true object size: in other words, because of size constancy. And we all notice in daily life that people do not shrink when they walk away from us; that is, there is phenomenological evidence to support the claim that size constancy is a basic tendency in visual perception.

Brunswik's outdoor observer could not, however, be given a variable stimulus with which to match real distant objects: this would have had to be extendable to 20 metres or more. So she gave a verbal estimate of perceived size. But is this what we mean by size constancy? To the author, a high-flying 747 looks very small indeed: his size constancy is clearly breaking down when looking upward through empty space. But asked to judge the size of the aircraft (rather than set a retinal match on a variable display) he would reply that it was about 50 metres long. This could be taken as evidence of size constancy, but it is not. It does not agree with phenomenal experience. It seems a pity that someone with Brunswik's research aims did not think this particular problem through more thoroughly. In fact, a convincing account of the basis of size constancy, this very important aspect of vision, did not emerge until 20 years after Brunswik's death.

Grouping and spatial proximity

This study is a rarity in psychological research in that it did not employ an experimental observer.

An important principle in Gestalt psychology is that of nearness or proximity. Wertheimer's (1923) classic demonstration revealed that stimuli arranged like those in Figure 2.5, in the previous chapter, become organized patterns. For example, it is conceivable that Figure 2.5a might be seen as a widely spaced inner pair of lines flanked by additional lines, or as four independent vertical lines. But neither of these organizations occurs: the figure is seen as two adjacent pairs of lines.

The Gestalt theorists explained this powerful tendency to group elements according to their proximity by postulating underlying fields in the brain that follow principles of attraction and repulsion. Thus, grouping is seen as a basic property of experience caused by lawful brain processes.

Brunswik's novel question, arising out of his functionalist approach, was: might grouping by proximity occur because it has survival value; is it the case that in the real world adjacent parallel lines tend to be associated by forming the boundaries of objects? The question can be answered by a simple analysis of parallel lines in a sample of the environment.

As an approximation to a representative sample of the 'existing ecology', Brunswik and Kamiya (1953) obtained several stills from a popular motion picture. The number of adjacent straight (or nearly straight), parallel (up to a deviation of 5°) pairs of lines was counted and their separations in the photograph measured. The pairs of lines were then classified by what they represented in the photographed scenes.

The results of this study are quite revealing. The (geometric) mean distance in the photographs between pairs of lines common to actual objects was 1.2 mm; that between lines representing 'ornamental divisions' (i.e., regular markings on surfaces) was 1.3 mm; lines delineating holes, gaps or spaces between objects had a mean separation of 2.8 mm. A correlational analysis restricted to separations of lines representing objects and those representing spaces between them yielded a coefficient of +0.34.This significant result confirms Brunswik's guess: *proximity has ecological validity*.

Wertheimer's discovery can now be seen in a new light: the law has functional value. Grouping is useful because it will commonly lead to the delineation of objects. It works because of the way the world is. A valuable insight has been gained by examining the relationship between perception and the environment in which it takes place. This is one of Brunswik's most original and successful contributions. Notice, too, how this pioneering experiment foreshadowed some of the modern examinations of Gestalt ideas that we described in Chapter 2.

An evaluation of probabilistic functionalism

Brunswik's approach to perception has been presented as clearly and convincingly as possible. If this attempt has been successful, the reader may agree that Brunswik had some stimulating and novel ideas, ideas which are the more impressive when one considers how long ago they were formulated. Why, then, has he had so little subsequent influence on theory and research? Those who knew Brunswik testify to his originality and cleverness, and yet his influence has been slight. A number of factors seem to have led to this state of affairs; the main ones are listed below.

Brunswik's style

Brunswik's first languages were Hungarian and German. His English, while always correct, is often difficult and even turgid. Compare the opening of Köhler's highly influential *Gestalt Psychology* (1947), which was written by one whose first language was German, with the ending of Brunswik's best-known (1956) work, *Perception and the Representative Design of Experiments*.

> There seems to be a single starting point for psychology, exactly as for all the other sciences: the world as we find it, naively and uncritically.
>
> (Köhler, 1947)

> Perception, then, emerges as that relatively primitive partly autonomous, institutionalized, ratiomorphic subsystem of cognition which achieves prompt and richly detailed orientation habitually concerning the vitally relevant, most distal aspects of the environment on the basis of mutually vicarious, relatively restricted and stereotyped, insufficient evidence in uncertainty-geared interaction and compromise, seemingly following the highest probability and smallness of error at the expense of the highest frequency of precision.
>
> (Brunswik, 1956)

The Brunswik paragraph is in fact a remarkable summary of an original theory, achieved in 68 words. And it is hoped that anyone who has worked through the present chapter will find it entirely comprehensible. It is, however, a hellish sentence. The style of Brunswik's English would matter less had he been more considerate in his reporting of experiments. Several diagrams in his publications are all but incomprehensible. It is commonly very hard to know what exactly happened in one of his experiments. The choice of symbols is often unfortunate and sometimes quite surreal. In one case, 'U-variables' are so named because 'u' is a vowel in the middle of the word 'population'; 'S' represents the environment in the lens model, 'U' now standing for individual differences. For every reader who learned to cope with this sort of thing, there must have been dozens who decided that Brunswik was not for them.

Brunswik's views on experimental design

Brunswik was opposed to 'classical psychophysics', in which all variables save one are controlled. The history of perception suggests, however, that such designs can be very fruitful ways of discovering the laws of perception. For example, much of what we know about colour vision has come from experiments using carefully controlled beams of monochromatic light; major researches into hearing have used only pure tones. We have not discussed this, but Brunswik was interested in the perception of faces and ran a multivariate experiment using schematic face patterns. His complicated design and the failure of the subsequent statistical analysis to reveal anything of real importance is in marked contrast to the much later work of Hess (1965, 1975), who showed that simply enlarging the pupil in a photograph of a face makes that face seem more attractive, even when the alteration remains unnoticed. This very simple (classical) experiment yielded a result as intriguing as any of Brunswik's in this area.

Brunswik's experiments

Brunswik often strayed from the ideal course he advocated for perceptual research. Consider his experiments, described earlier, which supported the idea of perceptual compromise. What, we may ask, is representative about an ensemble of Turkish coins or simple elongated rectangles? Brunswik's study of the phenomenon of grouping by proximity was a good idea and provided a critical test of the Gestalt explanation of the phenomenon, and hence of a whole aspect of Gestalt theory. The aim was clear: look at the disposition of adjacent lines in the world and see whether there is a functional basis for the grouping tendency in perception. But how representative were the photographs? They were not from the real world but from a film studio. They were pictures of constructed film sets. It is not necessary to labour the point, but anyone who reads Brunswik's experimental work after learning his views on the importance of ecological sampling will be surprised by the frequently contrived and artificial nature of his visual displays.

Some general criticisms

We shall not offer an exhaustive examination of all aspects of Brunswik's work. Interested readers may consult the Hammond (1966) reference previously cited. But it is proper to ask to what extent Brunswik's programme for perceptual research is a feasible one.

Brunswik emphasized the need to sample the environment or ecology of the organism. But here we meet a major difficulty: what is 'the' ecology? Is outdoors and indoors part of the same niche? We evolved in one and came to inhabit the other, so should we perceive them in a common way? Brunswik is silent on this point and it is clear that he greatly underestimated the problems associated with defining 'representativeness'.

The idea of the perceiver as intuitive statistician is one of Brunswik's most important assumptions and one that he shared with many later perceptual theorists. Perception involves making the best bet from imperfect information. Cues are weighted according to previous success or failure, a claim that emphasizes the role of learning. Of course, certain cues, such as pain-inducing stimuli, might be responded to reflexively, but generally the weighting of cues must depend upon experience.

Brunswik wrote at a time when much of American academic psychology was engaged by the problems of animal learning, and the 1940s produced several major theories in this area. At the same time, workers such as Hebb (1949) were stressing the role of learning in perception. Not surprisingly, Brunswik's theorizing reveals the influence of this Zeitgeist. However, subsequent research has shown that organisms, including humans, are surprisingly capable perceivers very soon after birth (evidence for this was reviewed in Chapter 2). If this shift of emphasis towards the innate aspects of perception continues, approaches such as Brunswik's will require important modifications (however, see the last section of this chapter).

Final remarks on Brunswik's theory

The desire to communicate complex ideas clearly and convincingly should be strong in any theorist who wishes to influence others. The neglect into which Brunswik's writings have fallen is partly his own fault. We wish to assert once again that Brunswik's view of perception is stimulating and original and that any reader who is now prepared to work through his writings will find that the ideas therein amply repay the effort.

Throughout this chapter we have maintained a fairly critical attitude towards Brunswik's work, particularly his empirical researches. Why, if there is so much to criticize in probabilistic functionalism, have we included this theory in the present book?

The answer to this question is that Brunswik should be valued less for what he achieved than for what he attempted. We believe that this was the first researcher to face up to the true complexity of perceptual processes, to recognize what a great achievement is represented by perceptual stability in an inherently uncertain world. The workings of our senses have been shaped by a successful evolutionary past, just as their structures have. And this shaping has been done by the complicated rich environment in which evolution took place: we must take this into account when thinking about perception.

Brunswik's assertion, that to simplify stimulus situations in the classical psychophysical manner was to ignore the properties of real-life stimulation, is convincing and well argued. The alternatives he offered – ecological sampling of stimuli, factorial designs, correlational assessment of performance – led to problems that he could not solve. But that does not detract from the originality of his ideas and the wisdom of his advice. And the complexity of his writing reflects not the confusion of a fool, but the vigorous efforts of someone who is trying to capture the complex truth as he sees it. Anyone who takes the trouble to read *Perception and the Representative Design of Psychological Experiments* will finish the book slightly puzzled but with a new and valuable perspective on perception.

The emphasis that Brunswik placed on the study of the ecology is re-emerging in contemporary work in perception. Workers who have adopted the direct perception paradigm have, as we shall show later, followed J. J. Gibson's lead in claiming that light and sound reaching the perceiver are rich in information. The task for the psychologist is to find within this richness invariant patterns that are capable of specifying a stable external world. Attention must be directed to the environment and its relationship to the perceiver. Indeed, Gibson's last book was entitled *The Ecological Approach to Visual Perception* (1979). And when we describe the computational approach to vision, we shall show that it is by carefully studying the environment that theorists can arrive at plausible constraints on their models of perceptual processes, a discipline which has been particularly fruitful.

At the start of this chapter, mention was made of the 'inference revolution' described by Gigerenzer and Murray (1987). Two facets of this revolution may be singled out: first, the idea of humans as statistical decision-makers – an idea of central importance in Brunswik's work; second, the incorporation into experimental psychology of the new statistical techniques pioneered earlier in the twentieth century by Fisher, Neyman, and Pearson.

It is important to stress at this point that the new statistical techniques were not adopted by psychologists as soon as they appeared. One route, which took some time, was via the work of Agricultural scientists in the USA. Only then did statistical testing start to enter experimental psychology.

The fact is that at the time when he was reporting the results of his experiments, Brunswik lacked advanced statistical competence – for the simple reason that all experimental psychologists did. The techniques for the analysis of results from multivariate experiments may have been known to some scientists; they were not part of the training of psychologists. Thinking about the lens model and the various correlational 'rays' running through it, suggests very strongly that one technique that could have helped Brunswik to handle his data more adequately is that of multiple regression. This is now taught to psychology undergraduates: Brunswik may never have heard of it.

We see, then, that the cliché 'ahead of his time' is the truth in Brunswik's case. He himself was insufficiently expert to manage the analyses of the complex experiments he wanted to do. And even had he been able to do so, the number of readers with the competence to understand and further develop his experimental programme would have been very small. In Chapter 2, we described the plight of some of the Gestalt psychologists who were forced to leave their home countries after the rise of the Nazi movement. Brunswik, too, was a refugee and he also had a short life. It is a cruel irony that by the time of his death academic psychologists were beginning to absorb the important ideas of inferential statistics and the mind, and were mastering the new statistical techniques that could have supported Brunswik's research programme and led to experiments as rich and complex as those Brunswik aspired to. In the final section of this chapter we shall describe one of these experiments.

A more recent development: the empirical theory of vision

Although they do not refer to Brunswik's theory very often in their publications, Dale Purves and R. B. Lotto and colleagues have published some recent work on visual perception that would have delighted Brunswik. Purves, Lotto, and colleagues at Duke University Medical Center and University College London have proposed an empirical theory of vision. The reader may wonder why this work is included in the present chapter, rather than on a later one on empiricism. We hope that by the end of this section, our decision to include it here will have become obvious.

Remember, from earlier sections in this chapter, that one of Brunswik's main concerns was to explain the relationship between things in the real world (3D distal stimuli) and the structure of the visual image (the 2D proximal stimulus). Brunswik, like others before him, knew that visual images contain information that is basically ambiguous. A particular contour in the visual image could have arisen from a small object close to the eye or a much larger one at a distance from the eye. Similarly, a trapezoidal shape in a retinal image could have arisen from an actual trapezoidal shape in the vertical plane or a rectangular shape tilted away from or towards the viewer.

This ambiguity is true not only in shape perception. A grey patch in the retinal image could have arisen from a surface of medium reflectance (a 'true' grey), or it could represent a highly reflective surface (a 'white') in shadow. Phenomena such as these are real problems for perceptual theorists.

There have been many attempts to explain how it is that our valid perception of real world events is achieved – what is the basis for this achievement? It will be remembered that Brunswik claimed that perceptual outcomes are based on 'best bets' concerning the nature of the world: this is essentially a statistical theory of vision (remember that red fruits are commonly ripe and edible, but some are dangerous; green fruits are commonly inedible, but some are edible).

The uncertain relationship between distal and proximal stimuli is at the core of the empirical theory. In Purves's and Lotto's own words (2003):

> The central hypothesis is that visual percepts are manifestations of the accumulated influence of visual experience with inherently ambiguous stimuli; therefore understanding what we see and why will depend on understanding the probabilistic relationship between stimuli and their sources that has shaped human visual physiology and its perceptual consequences.

This leads to the research question as to whether what we see accords with the probability distributions of possible real-world sources of visual stimuli.

As we as a species have survived, we are here because many thousands of our ancestors managed to resolve the complexities and ambiguities of vision. And they did so quickly: when a tiger is charging at one, or a spear is coming in on course, this is no time to think and reflect. It is important to get things right – form the correct perceptual decisions quickly, even at the expense of some small errors. The empirical theory states that through evolution and our personal experiences, we can resolve ambiguity by (unconscious) statistical analysis.

The reader will have noticed the similarities between the empirical theory and Brunswik's ideas. The following is an account of what we believe to be one of the most significant experiments published by this group of researchers. This is highly technical work and we shall have to simplify somewhat. But it is the sort of work that Brunswik could only have dreamed of, and it will become obvious why he could not even attempt it: the Purves and Lotto team have used modern techniques of high sophistication. The reference here is to Howe and Purves (2002). A much more extensive account can be found in Purves and Lotto (2003).

Howe and Purves (2002) carried out the following remarkable experiment. A laser range-finder was set up in two positions in the Duke University campus. The first was in the Sarah P. Duke gardens, the second in the nearby Duke forest.

This extraordinary range-finder has a range of 2–300 metres, with an overall accuracy of ± 25 millimetres. The range-finder was mounted at what is about the average human eye-height. Wide field images were acquired from the natural scenes. The result was that fine details of all objects included in each scene, together with accurate records of their 3D locations, were captured and stored in a computer. The next bit is rather complicated. An imaginary projection plane was placed at the polar origin of the stored data, which was the origin of the laser scanner. A region of the sampled 3D world was then projected (by computer, of course) onto an imaginary 2D plane, and this process was repeated after altering the orientation of the plane in 5° step-changes in azimuth and elevation. The result was a series of 15,000 2D slices (as it were) of the 3D world, each comprising approximately 196,000 pixels. It gets even more impressive: the 3D coordinates of each pixel in each 2D sample were then added to the database. It was as if a block of the real 3D world had been captured and brought into the laboratory for detailed scrutiny and analysis. Remarkable.

The next step was to analyse samples of the data as follows. Using powerful computer algorithms, all colinear segments in the images (indicating edges in the 3D world sample) were grouped together. The majority of these tended to lie close to the ground plane or towards the more vertical axes. As the authors say, '. . . fewer straight lines derive from leaves than tree trunks'.

It was now possible to look at the relationships between lengths in the 2D images (representing the retina) and all the colinear edges in the 3D images gathered from the laser scans of the environment. When this analysis was complete, the resulting frequency distributions showed certain distinct maxima. In other words, although it is *theoretically* possible that any contour in the retinal image could have arisen from an infinite number of 3D edges in the real world, in actual fact this is not true: there are non-random *patterns* in the information arriving at the retina. What the perceiver must do is to take advantage of these statistical patterns.

The work reported above will be very new to some readers. At this point we shall pause and offer a brief recapitulation. There are distal and proximal stimuli. The distal stimuli arise within the real world. Proximal stimuli are patterns of light energy arriving at the eye and forming images on the retina. The proximal stimuli are all we have to go on when perceiving the world. An age-old and fundamental question is how well proximal stimuli represent the distal world. The work by Purves, Lotto, and their colleagues attempts to answer this question by conducting meticulous quantitative analyses, of both the distal world and proximal stimulation. Their findings indicate that there are statistical patterns showing how well these two groups of stimuli correlate; in other words, how valid are patterns in proximal stimulation as guides to the external world?

As was stated earlier, it is unlikely that an individual perceiver could learn to utilize this patterned statistical information in a single lifetime (although experience could obviously refine the search). But in the psychoanalyst Jung's words, 'We are of an immense age'. The survival of our ancestors has led to much of this statistical skill being built into the visual system.

Although we have omitted certain technical details from this summary of the work of Howe and Purves, we have tried to stay close to the style of their thinking. We have not mentioned their interest in the horizontal–vertical illusion (the width of a square seems to be less than its height), a discussion of which actually forms the start of their paper, and an explanation of which ends their article. Neither has there been space to describe the many other applications of the empirical theory: the perception of lightness and brightness, colour perception, motion perception, binocular vision. In these areas also, it is beginning to look as though the central idea of the new empirical theory is yielding challenging insights.

Earlier in this chapter on Brunswik's theory, we quoted his assertion that perceivers must behave like 'intuitive statisticians'. He may have been right.

Endnotes

- Brunswik's *Perception and the Representative Design of Psychological Experiments* (1956) is a difficult book, but it contains the core of his ideas and is thus essential reading.

- Hammond (1966) is a useful source book, containing essays by a number of psychologists who were contemporaries of Brunswik. Some evaluate parts of Brunswik's theory and others attempt to relate his ideas to their own researches. The modification to the lens model shown in this chapter is explained more fully in Hammond's collection (see Chapters 2 and 3). Part 3 of the book is a reprint of some of Brunswik's papers. All Brunswik's publications are listed in an appendix.

- The following references are not cited directly in the text but may be useful in understanding Brunswik's approach: Brunswik (1938, 1939, 1948, 1955).

- Petrinovich (1979) contains valuable discussions of some of Brunswik's ideas. Brehmer (1984) attempts to show the relevance of Brunswik's approach to modern perceptual theory and research.

- Once again, the book by Gigerenzer and Murray (1987) may be strongly recommended for its description of the inference revolution and its comments on Brunswik's work.

- The work by Purves and Lotto and their colleagues described at the end of this chapter can be found in their book, *Why We See What We Do* (2003). The contents of this book cover a far wider range of topics than those we have described and is fascinating reading. The book is superbly illustrated.

4 The neurophysiological approach to visual perception

This chapter will describe some areas of perceptual research in which it has been suggested that psychological hypotheses can be replaced by known neural mechanisms. There are good reasons why a significant number of researchers in perception have always been in favour of this shift. In the first place, it is manifestly true that neural mechanisms underlie all behaviour. In an important sense *they* wrote these words and are now reading them. And there are those who believe that psychological knowledge is more secure when it can be linked to known physical structures. For example, the acuity of the eye falls off dramatically as one moves away from the central (foveal) region. The function linking falling acuity with degree of eccentricity is known sufficiently precisely to allow the prediction of visual performance in the periphery. But the reason why this falling-off takes place is now known: it is because of the increased ratio of rod to cone cells in the peripheral retina. The high degree of connectedness of the rod system results in high sensitivity through summation of outputs, but the price for this is the lowered resolution of the system. For many this is satisfying knowledge. A final reason for preferring neurophysiological explanations is simply that some researchers find it easier and more satisfying to think in terms of neural mechanisms, rather than in more abstract psychological terms.

The fact that perception, memory, and thought are all mediated by the central nervous system does not, however, force us to accept reductionism. The neural structures underlying mental events may be interacting in ways of which we cannot conceive and which could never be described using only the language of neurophysiology. This is said simply to warn the reader against too ready an acceptance of some of the claims to be outlined later in this chapter.

The approaches to be described have one thing in common: they invoke neural mechanisms in explanations of perceptual phenomena. We have of course met such an approach in the earlier chapter on the Gestalt theory. But this modern work differs from the Gestalt approach in two important ways. First, Köhler's physiology was highly speculative and, as it happens, largely incorrect; modern discoveries and theories are much more securely based. Second, the Gestalt psychologists, as phenomenologists, wanted to explain the richness of everyday perception; modern neurophysiological theories of perception are usually more modest in their aims. Typically, what they try to explain are basic sensory discriminations: how perceivers process some of the basic information contained in, say, the visual image, how this is coded and in what form it is sent onwards into the higher regions of the visual pathways. Such questions are very different from asking, for example, how familiarity affects our perception of objects, or why blue is almost certainly the world's favourite colour.

Once again, we shall follow the method of outlining some early pioneering studies in detail, before outlining some more recent developments. The reader should remember that only a small part of neurophysiological research is aimed at understanding visual processes. Much of the financial support for this type of research is, quite properly, directed to the search for the causes of brain disorder and the discovery of possible treatments.

The remainder of this chapter will cover the following topics:

- An outline of neural function.
- Three examples of the neurophysiological approach to perceptual theory: (1) colour vision; (2) feature detectors in the visual system; and (3) the visual system's responses to spatial frequencies.
- Classical computer models vs. parallel distributed (connectionist) networks.
- Two more recent technical developments.
- Some problems with the neurophysiological approach to perception.

An outline of neural function

None of the theoretical work to be described would have been possible without the remarkable gains in the understanding of the nervous system that have been achieved during the past 150 years. This is clearly not the place to undertake a history of neuroanatomy and neurophysiology, although this is a fascinating story and well worth reading. However, for those readers who are unfamiliar with neural structure and physiology and who lack easy access to specialist libraries, the following very brief treatment may be of some help. Other readers will skip the next section. Those wishing to learn more about the workings of the nervous system should consult the basic references given in the Endnotes to this chapter.

The nervous system

The term 'nerve' is used somewhat loosely. Major nerves are in fact bundles of nerve fibres: each human optic nerve, for instance, actually comprises approximately one million separate fibres. But 'nerve' is sometimes used to describe the basic unit of the nervous system, the neuron.

Neurons are specialized cells having a variety of shapes and sizes.

Basically, each neuron comprises a cell body with a nucleus, a complex arrangement of branching structures or dendrites, and one or more long processes or axons which run either to other neurons or to muscles or glands. The neuron receives stimulation via its dendrites and passes on stimulation via the axon. A typical neuron is shown in diagrammatic form in Figure 4.1

The connection between two neurons in a sequence is not physically direct. Activity in the first leads to temporary changes in a minute gap or *synapse* between neurons. Whether or not the stimulation from the first neuron is passed on depends upon the strength and timing of the changes at the synapse. Many neuronal endings may terminate on the dendrites or cell body of a single neuron. When a dendrite receives sufficient stimulation, the characteristics of its membrane at a local site suddenly change. As a result of rapid chemical processes, the permeability of the membrane alters in such a way that ions pass into and out of the cell. This results in a wave of electrical disturbance, or depolarization, which spreads away from the site of stimulation. This wave of electrical disturbance is decremental, tending to diminish with distance. But if several stimulating events occur within a short time or within a small area the electrical wave may be strong enough to reach the site where axonal conduction begins (the axon hillock).

Once a disturbance reaches the axon it is propagated according to a different principle: now the wave of electrical conduction is no longer decremental but all-or-none. That is to say, if an impulse begins to run along an axon it will continue to the end. And the size of the impulse is independent of the strength of the original disturbance – just as the speed of a bullet is independent of the strength of the trigger pull – provided this exceeds the threshold of firing. Neurons tend to code strength of stimulation as frequency; the stronger the stimulus, the more impulses per second. From the end of the neuron, activity spreads into the next synapse and, of course, this can lead to graded stimulation of the next neuron. Thus, the rapid all-or-none conduction down an axon fibre can be seen as a means of conveying graded information, translated into a frequency code, to another site in the body.

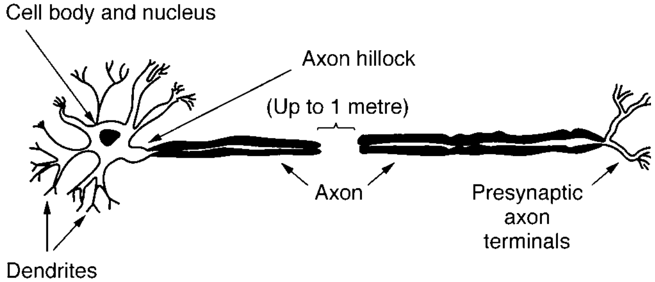
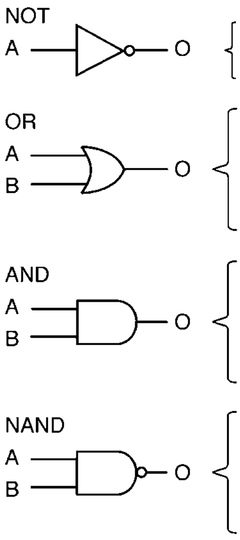


Figure 4.1 Diagram of a typical bipolar neuron.

The preceding account of the causal sequence between adjacent neurons has omitted an important phenomenon. We have described the pattern by which one neuron excites another, increasing the chance of the latter's firing. But neurons can also interact in an *inhibitory* manner. Thus, one neuron's activity, rather than inducing a wave of depolarization in another neuron, may actually cause a *hyperpolarization* of the next membrane, thus lowering the probability that the second neuron will fire. When we put these basic facts together, we can see that the activity between successive neurons affords: (1) threshold effects resulting from summation over space and time; and (2) positive (excitatory) and negative (inhibitory) interactions between these basic units of the nervous system. These facts are very significant, for it means that groups of neurons can behave in ways directly analogous to *logical gates* (see Figure 4.2)

Neurons as logical gates

In engineering terminology, an AND gate, for example, is a switch-like device which produces an output only when both its inputs are positive. An AND–NOT gate will not give an output if both inputs are simultaneously



INPUT		OUTPUT
A	B	O
ON	–	OFF
OFF	–	ON
OFF	OFF	OFF
OFF	ON	ON
ON	OFF	ON
ON	ON	ON
OFF	OFF	OFF
OFF	ON	OFF
ON	OFF	OFF
ON	ON	ON
OFF	OFF	ON
OFF	ON	ON
ON	OFF	ON
ON	ON	OFF

Figure 4.2 Four common logical gates. The function of the OR gate, for example, is to give an output if either or both its inputs are active. The AND gate gives an output only if both its inputs are active. Combinations of various gates can be used to build computing devices.

positive; an EXCLUSIVE–OR gate will give an output if either of two inputs occurs, but not if both occur. With such simple switching devices it is possible to build elaborate logical networks.[1] It is now possible to link these facts and draw an exciting conclusion:

(1) Neurons interact in various ways. One neuron can excite or inhibit another, increasing or decreasing the chance that the latter will fire.
(2) Logical gates are switches and can be used to build computing devices.
(3) Because of the ways in which they interact, neurons can simulate logical gates.
(4) Therefore, neurons can do something akin to computing.

This is a very important development in the history of neurophysiology. Some of the implications of this idea will be dealt with later in this chapter and in Chapter 7 on the computational theory of vision.

Neural processes take time. The discovery by Helmholtz in 1850 that the speed of conduction along a sensory nerve is in the order of 100 metres/second was vitally important. Neurons do not conduct instantaneously (as some had believed); they are relatively slow. ‘The speed of thought’ is not instantaneous, but is commonly slow enough to be measured, as is the speed of perceptual processes. Were this not so, psychologists would not have been able to discover nearly as much as they have about perceptual processes, and neurophysiologists would have had a much harder time trying to understand the ways in which neurons respond and interact.

Responding to change

This necessarily brief review of neural action may be completed by stating a final major principle that researchers have discovered: a large proportion of the various sensory neurons seem to have evolved to deal with *change* (see Figure 4.3). In numerous regions of the afferent nervous system, it has been found that the onset or offset of stimulation produces a rapid and marked increase in neural firing. But should the stimulation continue then, typically, the neural response returns to a value close to the resting baseline. We know that change is important in vision: if one looks into a completely homogeneous volume, for example if the head is placed inside an illuminated white sphere (known as a Ganzfeld), then vision will fade within seconds – the surface of the sphere softens to a fog, and eventually the sensation of seeing is lost completely (interested readers may experience the Ganzfeld effect by placing half a table tennis ball over each eye and looking towards a source of illumination). A similar fading occurs when the eye is effectively

---

[1] Philosophically trained readers will recognize the connection between logical gates and truth tables.

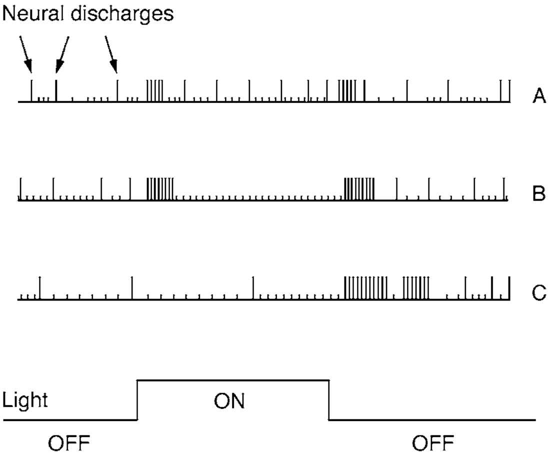


Figure 4.3 Three types of response from neurons in the optic nerve of the frog. Each spike represents a single neural discharge. Fibre A responds to light onset and maintains a steady discharge rate. Fibre B responds maximally to light onset and offset. Fibre C responds maximally to light offset. The combined effects of these responses results in the visual system responding most strongly to changing illumination. (Diagrammatic, after various authors.)

prevented from moving: this is known as the stabilized image phenomenon. Analogous effects occur in touch: an object placed on the skin is felt very clearly at first, but within a few seconds the tactile impression fades. These effects seem to reflect an underlying basic principle of economy of response. Change is always potentially important and there is an obvious evolutionary advantage in concentrating neural resources so as to maximize responses to it.

The question asked in this chapter is, how far can the knowledge such as that outlined above be employed in the solution of perceptual problems? Can one by-pass psychological theory and go straight to a causal, mechanistic account of sensory and perceptual phenomena? There are those who believe that this may be possible and it is to their work that we now turn.

Three examples of the neurophysiological approach to perceptual theory

To date, neurophysiological explanations in visual perception have been of various kinds. We shall draw examples from three areas of research: (1) the direct substitution of known neural/physiological mechanisms for hypothetical constructs; (2) the discovery of neural feature detectors; and (3) work that reinforces the growing belief that structures in the visual system can perform elaborate syntheses, as well as analyses, of incoming sensory data. This three-fold classification is somewhat arbitrary, but it does provide a structure within which to describe a selection of modern researches (later, we shall describe a computing system inspired by neural behaviour: the parallel distributed network).

Although elegant neurophysiological work has been done in other sense modalities, all the following examples are drawn from visual studies. The neural region of the eye, the retina, is actually an outgrowth of the brain and is thus a region of formidable complexity. We should not be surprised by some of the extraordinary processes that neurophysiologists have discovered there in the past few decades.

Neurophysiology and colour vision

The following examples of this substitution of known mechanism for psychological hypotheses are both drawn from the area of colour vision. They are good illustrations of the successful application of neurophysiological knowledge to classical psychological problems.

The two most important facts about colour vision in humans (and some other species) are, first, that our colour vision is trichromatic, and second, that we experience highly predictable contrast and fatigue effects.

The trichromacy of human colour vision means simply this: suitable mixtures of three wavelengths of light can match most of the hues that a person is capable of perceiving. These primary wavelengths need not be precisely specified, provided that: (1) they span the visible spectrum – there is a wide range of choices among the blues, greens, and reds; and (b) no two primaries should be exactly complementary (for each hue in the visible spectrum there is another, complementary hue which, when added to the first, yields an achromatic mixture; such complementary pairs must be avoided when choosing the primaries).

Most people are greatly surprised when their trichromacy is first demonstrated to them. It is a memorable experience, particularly when it is seen that equal amounts of the three primaries mix to produce white: as the intensity of the third light is increased, all colour simply fades away. Moreover, the matches made are highly stable. For example, if one produces yellow by adding red and green light, the yellow can be made indistinguishable from that seen in the ‘yellow’ portion of the spectrum (wavelengths of approximately 560–580 nm). If one then biases colour perception by fatiguing the eye with, say, orange light, *both yellows change in exactly the same way*. The stability of the match is also maintained when an additional coloured light is added to both yellows.

Colour contrast and fatigue effects are equally remarkable phenomena. If a red square is placed on a grey ground and fixated for a few moments, one comes to see a greenish tinge surrounding the red. An intense green light induces a reddish after-image; blue light induces yellow, and vice versa (note that red and green and blue and yellow are complementary hues in that they mix to form neutral greys). Anyone can experience these effects by simply staring at a coloured light (*not the sun*) for a few moments and then looking at a white surface. A related phenomenon may be observed on brightly lit snowscapes, where it can be seen that shadows are blue because the eye adapts to the yellow sunlight and sees the snow as white; when the sunlight is interrupted by an object to form a shadow, the complementary blue appears (this was first brought to general attention by the Impressionists)[2].

The Young–Helmholtz theory

In the nineteenth century these very reliable and interesting phenomena gave rise to a number of theories of colour vision. The first, known now as the Young–Helmholtz three factor theory (Helmholtz, 1909–1911, trans. 1924–1925), attempted to explain trichromacy as follows. Suppose that the eye contains three types of receptor, each maximally sensitive to a portion of the spectrum (Young originally proposed three pigments, Helmholtz three types of retinal cones; historically, the difference is not important). Then, if the eye is illuminated by a particular hue, the type of cell whose sensitivity is closest to the wavelength of the hue will fire strongly, while other types of receptor will respond less vigorously. Yellow light will stimulate the cells sensitive to the red and green parts of the spectrum about equally. White light will stimulate all three types of receptor, evoking the achromatic response (see Figure 4.4).

The Young–Helmholtz theory of colour vision does have weaknesses. It does not readily explain the stability of yellow, a hue which is still seen in intensely strong light when all other hues apart from blue vanish: how can the yellow remain when the contributing receptors (in the red and green regions) do not appear to be functioning? Also, the theory has some difficulty over certain forms of colour vision deficiency. Finally, the theory does not account for the changes in perceived hue that accompany changes in the intensity of coloured stimuli (the Bezold–Brücke effect). Nevertheless, the Young–Helmholtz three-factor theory has proved to be very useful and durable. It is the most widely cited theory in the history of colour vision research.[3]

Hering's opponent process theory

The major rival to the Young–Helmholtz theory at that time was Hering's opponent process theory (Hering, 1890). This theory postulated the existence in the optic nerve of three processes capable of functioning in, as it

---

[2] The blueing effect is not simply a contrast phenomenon: it owes something to the differential absorption properties of snow.

[3] Interestingly, the Young–Helmholtz hypothetical primary colours are those used in modern television sets.

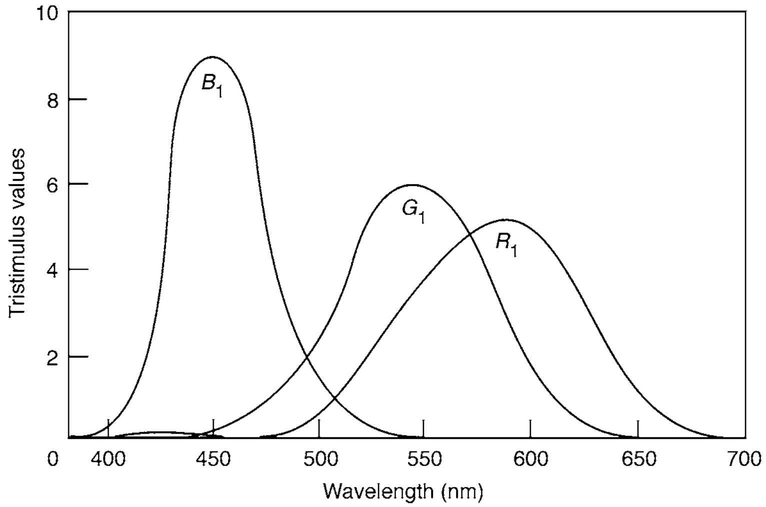


Figure 4.4 A diagrammatic illustration of a three-component theory of colour vision. B_1, G_1, R_1 represent the hypothetical receptors postulated by the Young–Helmholtz theory. The vertical axis can be interpreted as the relative absorption efficiency of each receptor as a function of wavelength. (First described by Helmholtz, 1909–1911.)

were, opposite directions. In the ‘anabolic’ direction, the processes give rise to the sensations of red, yellow and white; in the ‘catabolic’ direction, these same processes give rise, respectively, to green, blue and black (see Figure 4.5). Thus, the phenomenon by which complementary hues mix to grey is accounted for in terms of a balanced neutral point in the relevant opponent process. Fatigue and contrast effects are handled just as easily, as is the fact that one cannot see, for example, blue and yellow at the same time in the same place. And while there are blueish-greens, there is no blue–yellow sensation, neither are there any reddish-greens or blackish-whites.

Hering’s opponent process theory also has its weaknesses. For example, it predicts a form of yellow–blue colour blindness that has never been found, and the theory does not yield an entirely satisfactory account of the brightness of colours. Nevertheless, it provides an explanation of some very important phenomena. The debate between these two very different theories (or their more recent counterparts) has been lengthy. And one can see why: each explains some of the facts of colour vision but not others. But the theories are very different: how could they both be right?

We now know that both the Young–Helmholtz and the Hering theories are essentially correct, within limits. Our confidence that this is the case is one of the triumphs of visual research. An account of the work that confirmed both three-factor and opponent-process theories will demonstrate the way in which actual neural mechanisms can displace hypothetical constructs.

Human (and animal) data from colour vision experiments have led to the

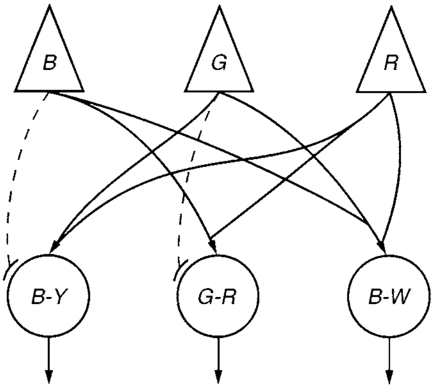


Figure 4.5 A representation of a three-pigment/opponent-process colour vision model. Three types of receptor (*B*,*G*,*R*) responding to short-, medium- and long-wavelength light send outputs to opponent cells (*B–Y*, *G–R*, *B–W*). These outputs may be excitatory (solid lines), or inhibitory (dashed lines). For example, the *B–Y* unit receives an inhibitory from the short-wave receptor, resulting in a 'Blue' output from the cell. Yellow light will stimulate the *G* and *R* receptors equally; they excite the *B–Y* unit which then produces a 'Yellow' output. (Constructed from the writings of several authors.)

construction of quantitative models of colour vision in which various triads of hypothetical pigments are evaluated. By constructing absorption curves for these hypothetical pigments, one can test whether it is possible to account for various aspects of colour performance – particularly matching tasks, the colour confusions made by colour-deficient judges, and the relationships between hue and other aspects of colour, such as lightness, brightness, and saturation. As a result of many years of careful measurement, there are now good data that can be used to predict various colour phenomena. However, the hypothetical pigments in this research are selected to give the best fit to the performance data: there is no direct evidence that pigments in the eye exactly match them. How satisfying it would be to locate the real pigments and to know once and for all what underlies human trichromacy. This is a goal that has eluded visual researchers for many years.

No single researcher can be given the credit for finding the actual three cone pigments. However, many would agree that an important breakthrough came when Rushton (1964) perfected a technique that made it possible to search for pigments in the living eye.

In essence, *microspectrophotometry* entails shining a narrow beam of pure monochromatic light onto the cells of the retina, trapping the returning beam and measuring the difference between the two. In this way it is possible to assess the absorption properties of retinal cells. Described so baldly, the work sounds relatively simple; in fact perfecting the technique took many years of intensive research. Rushton's work was extended by MacNichol

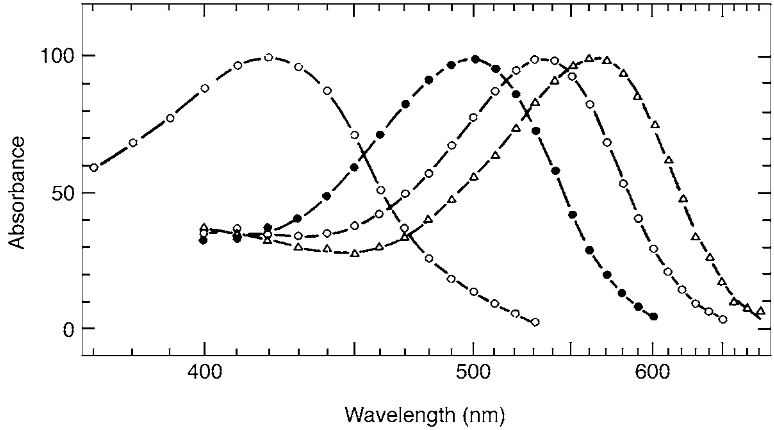


Figure 4.6 Actual absorption data obtained from isolated human cone cells. Microspectrophotometry has revealed the presence of three distinct cone pigments with absorption maxima at 420, 534 and 564 nm (the filled points are from the rod pigment, rhodopsin). The basis of trichromacy has been discovered. (From Dartnall, Bowmaker, & Mollon, 1983. Reproduced with permission of the Royal Society of London.)

(1964) and Dartnall, Bowmaker, and Mollon (1983), who worked with isolated cone cells. When these various researches were combined, it became certain that the cone cells of the retina do in fact contain three different pigments. Each has a wavelength to which it is maximally absorbent and the three peak sensitivities are at 420, 530, and 560 nm (see Figure 4.6 and compare it with Figure 4.4). This is a most satisfying result. It enables us to call cells containing the pigments the short-, medium- and long-wavelength colour receptors of the eye. This is exactly what is required by trichromatic theories, such as the Young–Helmholtz theory outlined earlier. Helmholtz was essentially correct (that he chose cone types, rather than pigments, does not matter). The basis of visual trichromacy has been discovered.

The story just told represents a remarkable gain in our knowledge of colour vision. It is not, however, an adequate explanation of all colour vision phenomena. The existence of a cone type containing a long-wave (red-absorbing) pigment does not by itself explain how we distinguish between different reds – how our colour discrimination is so good – neither does it readily explain those phenomena which prompted the opponent process theory outlined earlier.

The goal of finding opponent processes was achieved with the success of Svaetichin (1956), who discovered an electrical potential in cells of the fish retina which responds differentially to coloured light in the following manner: at short wavelengths the potential responds positively, at long wavelengths negatively (see Figure 4.7). Other cells in the retina produce a similarly selective response to blue and yellow light. Then De Valois (1960)

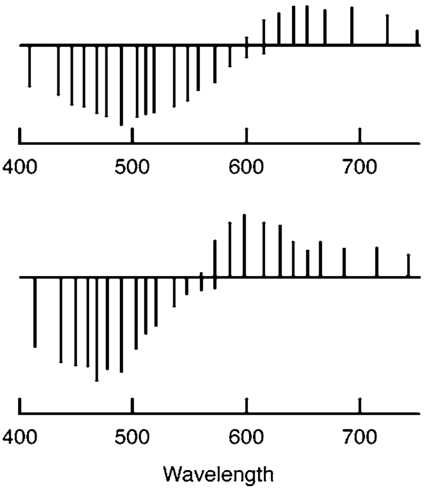


Figure 4.7 Svaetichin's discovery of opponent-process responses in cells of the fish retina. The slow electrical potential (the S-response) changes its polarity as a function of wavelength. In the upper part of this diagram the wavelength changes (from left to right) are from yellow to blue. In the lower part of the diagram the wavelength changes are from red to green. (This simplified diagram was constructed by Bettina Newman from data published by Svaetichin, 1956.)

found cells in the lateral geniculate nucleus (a relay station between the retina and the visual cortex) of the monkey that also respond in an opponent manner to wavelength. These cells respond by increasing their firing when the eye receives light from one end of the spectrum and decreasing their firing when the light is from the other end of the spectrum. Refinements in this research have now uncovered +Blue –Yellow, –Blue +Yellow, +Red –Green, and +Green –Red lateral geniculate cells, *all behaving in a manner suggested by opponent process theory.* Once again, neurophysiological work has demonstrated the essential correctness of an abstract theory of colour vision. It has also yielded a satisfying explanation as to why the theory works and unites two very different theories. That is to say, retinal cone cells in the eye do absorb light by the action of three pigments, and these three pigments underlie trichromacy. At the same time, cells in the visual pathways located inwards of the cone cells use the outputs of these cells and respond differentially to them, producing the sharpening postulated by opponent process theory.

Taken together, these two sets of findings justify the claim that theoretical constructs at this level of colour vision research can be replaced by known neurophysiological mechanisms. This in turn allows research to be directed toward new problems: how the various pigments and opponent-process cells are arranged in other species; what is lacking in those people who have impaired colour vision, and so on. Of course, the discoveries do not signal the end of colour vision research and theory. Many questions remain, in particular how we perceive coloured surfaces where texture and hue interact; how the 'true' colours of things can be perceived when illumination is changing; why some animals, such as the cat, have cone cells and visual pigments in the retina but find it difficult to learn colour discriminations. But at the basic sensory level of explaining trichromacy and contrast and fatigue effects, neurophysiology has given us definite answers. Small wonder that some believe this to be the eventual fate of many other perceptual phenomena.

Feature detectors in the visual system

As an example of another type of neurophysiological theorizing in perception, some modern research into the perception and recognition of shape will be described, with particular emphasis on the search for feature detectors.

Recognition of the importance of shape perception and discrimination came early in the history of psychology. Mach (1836–1916), aware that contours play an important role in delineating shapes, solved some of the psychophysical problems of contour extraction. He established that contours appear whenever a gradient of lightness or brightness changes suddenly (technically, this is the second differential of the intensity gradient). The Gestalt psychologists demonstrated how shapes emerge from the ground, they stressed the importance of shape constancy in perceptual stability, and they showed how priority appears to be given in perception to balanced, simple, symmetrical shapes, according to the law of Prägnanz.

By the early 1950s there were enough facts to fuel a theoretical controversy. Some workers, for example Hebb (1949), claimed that there was evidence to support a learning interpretation of shape perception. Hebb's theory assigned an important role to eye movements in the creation of 'cell assemblies' mediating subsequent shape recognition. But at the same time ethologists, studying animal behaviour under natural conditions, found evidence of innate recognition of certain shapes. For example, shapes comprising only large and small discs induce unlearned gaping responses in nestling thrushes (see Tinbergen, 1951) that are the same as those induced by parent birds. The ethological literature contains many other examples of this kind of innate responsiveness to shapes.

Another development in the 1950s was the advent of digital computers capable of restricted pattern recognition. This encouraged the development of psychological models of shape perception and recognition (generally subsumed under the heading 'pattern perception'). What eventually emerged were two main types of model: template matching and feature detection.

Template-matching models designed by Selfridge and Neisser (1960) and Uhr (1963) recognize patterns or shapes by noting their similarity to canonical forms. Such an idea is illustrated by those educational toys for infants, in which solid shapes can be posted only through the correct apertures in a box. However, such template-matching models have major flaws: how is it, for example, that we can recognize a certain letter when it is presented in an unusual typeface, or in a different size, or at a different retinal location? Our ability to do all these things presents serious problems for this type of model.

Feature detection models were designed to avoid the difficulties described above. They work by analysing shapes into component parts or features. For example, Selfridge's well-known pandemonium model (Selfridge, 1959) postulates peripheral, low-level feature detectors, many of which are triggered by shapes falling onto a receptor surface. The outputs of these detectors are weighed at higher levels of the system, with the detector (actually called a demon) that 'shouts' loudest having the best chance of its output being accepted for further processing. The shape finally arrived at is based on combinations of features detected by the demons.

In a different context, Sutherland (1957) claimed that those shape discriminations of which the octopus is capable could be explained by assuming that visual stimuli are analysed in terms of horizontal and vertical features, but not by oblique ones. In yet another context, human perception and thinking, Bruner (1957) suggested that patterns are examined for key attributes, which are then related to categories created by the perceiver beforehand. Hence, some cognitive activity is believed to precede actual shape recognition.

For now it suffices to say, first, that the importance of shape perception and recognition has long been recognized by psychologists; second, that this is still a live issue. For example, Marr, whose important work will be the subject of Chapter 7, stated quite explicitly that he was attempting to formulate a theory '. . . in which the main job of vision was to derive a representation of shape' (Marr, 1982).

These, then, are some of the theories that have arisen in response to the challenge of shape perception. We shall now attempt to show how such theorizing is being influenced by other discoveries in neurophysiology.

In a pioneering study of the responses of the nervous system to stimulation, Adrian (1928) found that tactile sensory fibres in the limb of a monkey respond whenever a region of skin is stimulated. Adrian coined the phrase *receptive field* to describe the relationship between a region of a sensory surface, such as the skin, and neural cells inwards of the surface that receive messages from it. The concept of the receptive field is now centrally important in visual neurophysiology.

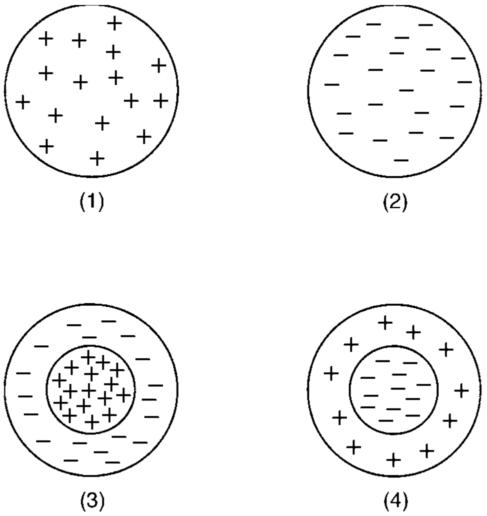
Adrian's work on the tactile receptive fields was quickly extended to other areas. From the work of such distinguished researchers as Hartline (1938, 1940), Barlow (1953), Kuffler (1953), Lettvin et al. (1959) and Maturana et al. (1960), knowledge of receptive fields grew rapidly. It was found, for example, that receptive field organization exists in the frog retina and in the optic nerve fibres of the cat (whose visual system shares many important characteristics with our own). Many visual receptive fields have a circular organization. In some of these fields central excitatory areas are surrounded by concentric inhibitory regions. The result is that stimulation in the centre of a visual area results in increased neural activity, but this can be inhibited by stimulation of the surrounding area – the on-centre/off-surround fields. The opposite organization is found in what are described as off-centre/on-surround fields. The quality of these researches was recognized in the award to Hartline and his colleagues of the 1967 Nobel Prize.

The possible relevance of this research for the psychology of perception was further demonstrated by the work of Lettvin et al. (1959) and Maturana et al. (1960) on the frog's visual system. Recordings from fibres in the optic nerve produced a very exciting discovery: the frog's visual system appears to respond in a very limited but selective manner to stimulation at the retina. Some cells produce a prolonged response to edges. Others respond when small dark objects are moved across the visual field (hence their name,'bug detectors'). There are cells that respond maximally to changes in contrast in the visual field. Others respond when their visual fields are darkened. Finally, there are cells that respond inversely to light intensity and thus appear to be dark detectors. These remarkable findings are doubly significant. Obviously they show that the visual world of the frog must be very different from our own. First it appears to be a simple world, restricted to those stimulus attributes that are vital to the frog's survival: the presence of small prey, the shadows of possible predators, the safety of darkness. Second, these aspects of the world that the frog must perceive in order to survive are extracted automatically. As we have seen above, neural mechanisms in the frog retina extract features from the visual image. That this processing is thus peripheral rather than central may be explained in part by the fact that the frog does not have a very complex brain – for example, it lacks a cortex. But warm-blooded vertebrates do have complex central nervous systems. When some of the researchers listed above turned their attention to the visual system of the cat (which, like many other warm-blooded vertebrates, has a well-developed visual cortex), they found that receptive fields can also be found in more central regions of the nervous system.

We shall now describe what has become a classic set of experiments. The work by Hubel and Wiesel (1962, 1977) to which we now turn has been described by some psychologists as the most important set of discoveries in the history of physiological psychology. The quality of this research was recognized in the award to the authors of the 1981 Nobel Prize.

Hubel and Wiesel (1962, 1977) succeeded in recording the electrical responses of single living cells in the visual cortex of the cat and the monkey to various patterns of stimulation. To appreciate the magnitude of this achievement, one must realize that cortical cells are microscopically small, so that to record from them without destroying the cells requires the use of exceedingly fine microelectrodes (these are so fine that the tip is invisible, even under a microscope). Then the electrode must be positioned very carefully in the cortex using precision stereotactic instruments. The aim is to make contact with the outer wall of the cell without puncturing and destroying it. The researchers must be certain that they are actually recording from a living cell – which, of course, they cannot see. Finally, the experimental animal must be kept alive under anaesthesia while the retina is stimulated in a controlled manner. It took researchers many years to overcome these formidable technical problems.

Hubel and Wiesel have described how one day, while trying vainly to induce a response in a cortical cell, they accidentally moved the slide in their projector so that the edge of the slide moved across the experimental animal's visual field. The cortical cell immediately responded to this moving edge. Subsequent experiments showed that what had been discovered were receptive field organizations in the cat's visual cortex. However, unlike the simple, circularly organized receptive fields found previously in the retina and lateral geniculate body, these cortical fields are thinner and more elongated in shape. They respond to the presence in the visual field of moving edges or contours having a particular orientation. Some cortical cells respond to vertical lines and their response falls off as the lines are changed away from the vertical. Other cells respond to horizontal or oblique lines and edges. Figure 4.8 summarizes some of these discoveries.



(1) (2)

(3) (4)

Figure 4.8 Examples of some of the types of receptive field organization in the vertebrate visual system. Each diagram represents an area of the retina monitored by a retinal ganglion cell. The signs represent the responses of the ganglion cell (+ = excited, – = inhibited) when light falls onto the receptive field. (1) and (2) represent the simplest forms of receptive field; (3) is an on-centre/off-surround field; (4) is an off-centre/on-surround field. Receptive fields are found in more central regions of the visual system. Not all fields have such clearly defined circular arrangements.

Not all the cells explored in the cortex by Hubel, Wiesel, and others have receptive fields, but subsequent studies of those that do uncovered some remarkable facts about the visual cortex of the cat and, later, the monkey. For example, some cells in the visual cortex exhibit a vertical, columnar organization. As one penetrates deeper into the cortex below a particular site, the column of cells produces subtly changing responses to stimuli, but all cells in the column exhibit a preference for the same orientation of the stimulus. Cortical cells show different types of responsiveness. In some the receptive field is an elongated area with excitatory and inhibitory regions. Others respond positively to appropriate stimulation anywhere in the relevant portion of the retina and have no inhibitory regions. Still other cells are indifferent to the orientation of the stimulus but respond selectively to patterns of a particular height and width. Many cortical cells are binocularly driven and can be induced to fire by stimulation of either eye.

The responsiveness of some cells is bizarrely specific: for example, Gross, Rocha-Miranda, and Bender (1972) found cells in the macaque monkey's inferotemporal cortex (a region of the brain at a distance from the visual cortex, which is implicated in certain forms of visual recognition) which respond selectively to the image of a hand.

It is important to stress at this point that the discovery, through microelectrode recording and other techniques, of a cell responding specifically to, for example, the image of a hand, is not interpreted by neurophysiologists as being the brain's only response to that hand. It is likely that adjacent cells, from which recordings are not being taken, are also responding to the image, but in these cases the responses may be to subtly different aspects of the hand. It is probably closer to the truth to conceive of groups of neurons acting in loose confederations, with each group showing some specialization (visual areas of the cortex vs. auditory areas, motion detectors vs. colour detectors, and so on) rather than any complete function being performed by a single specialized neuron.

Receptive fields are present at birth, a finding that gives some support to the nativist view of perception advanced by Gestalt psychologists (see Chapter 2). However, early experience can modify the nature of the fields. For example, Blakemore (1974) reared kittens in artificial environments comprising either vertical or horizontal striped surfaces. After varying periods of time in such environments the kittens were examined in two ways. First their ability to discriminate contours was tested. It was found that kittens reared in a vertical striped environment showed impaired acuity to horizontal stripes, and vice versa. The animals were not blind to the unfamiliar stripes, but their performance made it obvious that the stripes were not perceived as clearly as those in the familiar orientation. Second, when receptive fields were examined in the visual cortex of these animals, cells were found which respond normally to stripes in the familiar orientation, but the cortical responses to stripes in the other orientation were severely reduced. It was not that the cortex had actually lost a number of functional units, but rather that an abnormal number had developed to match the orientation of the striped rearing environment.

Thus, overt behaviour and the cellular responsiveness of the visual system both indicate that some early experience is needed for normal development of receptive fields, and that abnormal experiences can bias them. There is, moreover, a *critical period* during which experience is particularly important, and this exists between approximately 3 weeks and 3 months after birth. Similar conclusions emerge from variants on this work in which, for example, animals are reared with one eye permanently closed for a period to see the effects this has on binocularly driven cortical cells. Thus, this neurophysiological research seems to support a modified nativism, in which the elements of perceiving are present at birth, but not in a rigid or unmodifiable form.

Readers wishing to learn more about this important research will find excellent accounts in the source books listed in the Endnotes to this chapter. We shall now attempt to show how the discovery of receptive fields has influenced psychological theory.

One obvious interpretation of the discoveries by Hubel and Wiesel and subsequent researchers is that the feature detectors suggested by theories of shape perception have been found. Just as some had supposed, it seems that neurons in the brain (at least in cats and monkeys) are capable of responding selectively to certain aspects of stimuli. These can be simple features, such as lines in particular orientations; more complex relationships, such as particular lengths *and* widths; and very complex combinations of features, such as a hand shape. Small wonder that this research quickly attracted the attention of a great many psychologists.

Some of the best-known psychological and theoretical researches to follow Hubel and Wiesel's discovery were those of Julesz (1981). Julesz has investigated the properties of visual textures to see which can and cannot be effortlessly discriminated. Following a long series of investigations, Julesz claims that the basic building blocks of visual texture are dots, elongated blobs, and terminations of lines. And in describing the role that these *textons* play in perception, he makes specific reference to the work of Hubel and Wiesel – not surprisingly, for these are exactly the aspects of stimulation that their feature detectors can extract from images. Here, then, is clear evidence of what is known to occur in the visual system having a major influence in an important area of perceptual theory.

The emphasis so far in the two strands of research described above has been on neural *analysis* of sensory data – how information about colour and shape might be extracted from the visual image by simple neural mechanisms. We shall postpone further comment on these researches until we have given one more example of the impact of neurophysiology on perceptual theory. In this we shall attempt to explain how neural mechanisms might be capable of synthesis as well as analysis. The work to be described followed quite naturally from that above, and some of the researchers have worked in both areas.

Spatial frequencies

Spatial frequencies are associated with lines and edges, features that are of vital importance in visual perception. The concept of spatial frequency may be explained in terms of one of the widely used research tools in this area, the visual grating.

A grating is a display comprising alternate light and dark stripes. In visual research these stripes commonly do not have sharp edges but vary smoothly from light to dark and vice versa (Figure 4.9). Such *sinusoidal gratings* are used for technical reasons, in particular because they can be analysed by a powerful mathematical technique, Fourier analysis, that permits complex grating patterns to be decomposed into simpler sinusoidal components.

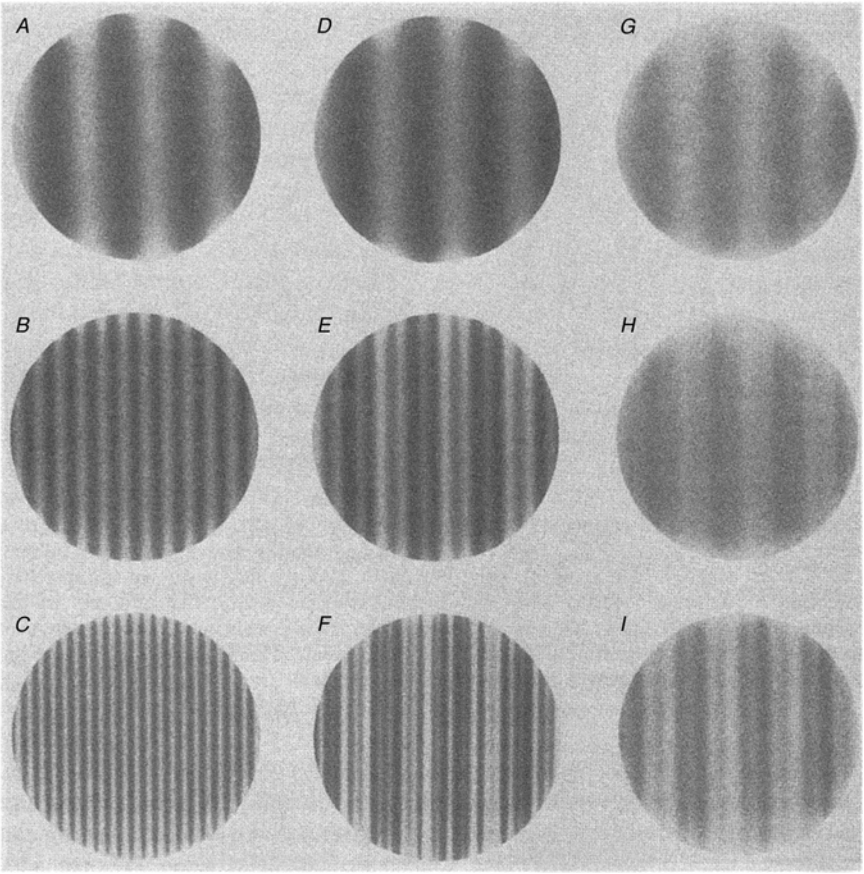


Figure 4.9 Simple and complex sinusoidal gratings. The top row of gratings differ only in contrast, those in the leftmost column differ only in spatial frequency. The remaining gratings are formed by combining the simple row and column gratings. (From Sekuler & Blake, 1985, with permission. Copyright © The McGraw-Hill Companies, Inc.)

Any grating can be described in terms of four independent properties:

(1) The *contrast* of a grating is simply the difference in brightness (or luminance or reflectance) between the light and dark areas; low-contrast gratings are harder to see, other things being equal. (The black–white contrast on this page is approximately 70% on a 0–100% scale.)
(2) The *spatial frequency* of a grating is a function of the width of the stripes, which in turn defines the number of alternations of light and dark across unit distance. A measure of spatial frequency is the number of changes per degree of visual angle.
(3) The *orientation* of a grating simply describes whether it is horizontal, vertical or oblique.
(4) The *phase* of a grating is taken from any arbitrary starting point: is the stripe in that position light or dark?

Simple gratings are those formed from a single spatial frequency. Complex gratings (Figure 4.9) are formed by adding simple gratings. Conversely, complex sinusoidal gratings can be analysed into their basic components.

It is obvious that one can have gratings with stripes that cannot be seen, either because the stripes are too fine to be resolved, or because the contrast between the light and dark areas is too low. It follows that there are two distinct thresholds associated with the detection of a grating.

In what has become a classic study on spatial frequency detection, Campbell and Robson (1968) measured these two thresholds in human observers. Using electronically generated sinusoidal gratings, Campbell and Robson selected a particular spatial frequency and set the contrast so low that the grating lines could not be seen. The contrast was then raised until the stripes were just visible. Then the spatial frequency was changed and the process repeated. Campbell and Robson presented their threshold data in a new form of graph, the contrast sensitivity function, which has provided valuable insights into the process of seeing. As Figure 4.10 shows, the interrelation between threshold contrast and resolution of spatial frequency takes the form of a curve. This curve is an exceptionally useful way of describing visual performance. It predicts the fineness of detail that can be seen at particular contrast levels; it is the best way yet of comparing the vision of different observers, and it allows us to compare human vision with that in other species. Note in Figure 4.10 that the cat is very sensitive to low spatial frequencies. This means that cats are able to see faint shadows that we cannot, which might explain the age-old association of cats with supernatural phenomena. The contrast sensitivity function has been described as ‘a window of visibility’. Interestingly, although it has long been known that other species can rival or even out-perform humans on traditional measures of acuity (the hawks have better resolving power) and on traditional measures of sensitivity (some nocturnal creatures have very high sensitivity), the human eye has the best all-round performance in terms of the contrast

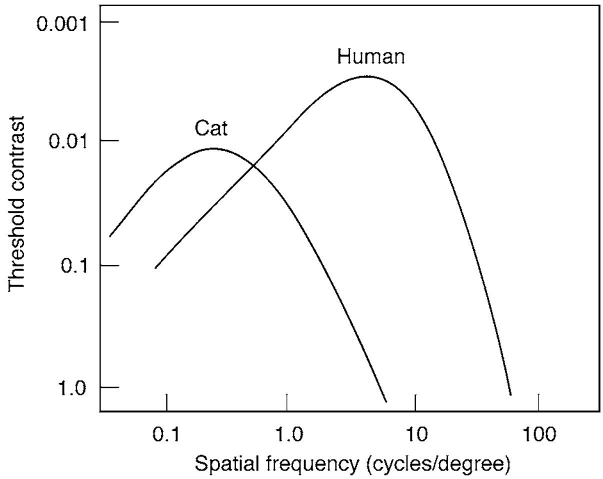


Figure 4.10 Contrast sensitivity functions. The contrast of a grating is defined as a ratio: $(l_{max} - l_{min})/(l_{max} + l_{min})$ where l_{max} and l_{min} are the intensities of the lightest and darkest regions of a grating. The contrast sensitivity functions are obtained by selecting a particular spatial frequency grating and increasing the intensity of the lighter regions until the grating is just detectable. This is continued over a range of spatial frequencies. In this figure, both axes are plotted on logarithmic scales. Note that although the human function is generally superior to the cat's, the cat is more sensitive to low spatial frequencies.

sensitivity function: it has the largest area under the curve, the largest window.

The use of gratings heralded a new approach to the measurement of visual performance, an approach which was to yield important new theoretical insights into the process of seeing, and which quickly led to the discovery of new and important phenomena. Three examples will illustrate the intriguing nature of these discoveries.

Spatial frequency channels

Campbell and Robson (1968) investigated the perception of complex gratings in a series of threshold determinations. Remember that a complex grating can be formed by adding a series of simple sinusoidal gratings (Figure 4.9). However, when one looks at such a grating, the components are not obvious – they do not appear in consciousness – and in a sense they cannot be perceived. But when Campbell and Robson examined the contrast sensitivity functions for complex gratings they made an interesting discovery. With the display initially appearing a uniform grey, the contrast of a grating is raised until the observer becomes able to detect the spatial frequency to which he or she is most sensitive. Further increases in contrast reveal the presence, one by one, of the other spatial frequencies contained in the complex grating. And each threshold is the same as it would be if that particular spatial frequency had been presented in isolation. This is a most intriguing finding, particularly when one looks again at Figure 4.9 to remind oneself that the component frequencies are not perceptually distinguishable.

Later, Blakemore and Campbell (1969) discovered an interesting adaptation phenomenon. When a subject fixates a particular grating for a period of time and then has his or her contrast sensitivity function assessed, a drop in sensitivity is observed. Such fatigue effects are common in vision and this one was not surprising. However, the strange thing is that the effect is not general, but is limited to those spatial frequencies close to that of the adapting grating. Similarly, fixating a horizontal grating reduces sensitivity to nearby frequencies, but only when these are horizontally arranged; there is no loss of sensitivity to vertical gratings. In both cases fixating has presumably fatigued some process, but the process is not general: it is orientation- and frequency-specific.

As a final example, it has been found that fatigue/bias effects in the perception of spatial frequencies are not limited to threshold stimuli. Fixate the left half of Figure 4.11 and then look between the two right-hand gratings. It will be found that the apparent spacings of these two (identical) gratings will have changed, a suprathreshold effect first reported by Blakemore and Sutton (1969).

These discoveries provoked an exciting idea: *the visual system conveys information about spatial frequencies in tuned channels*. This was an insight which was to have a considerable impact upon subsequent theorizing about the visual system, as will be shown in Chapter 7.

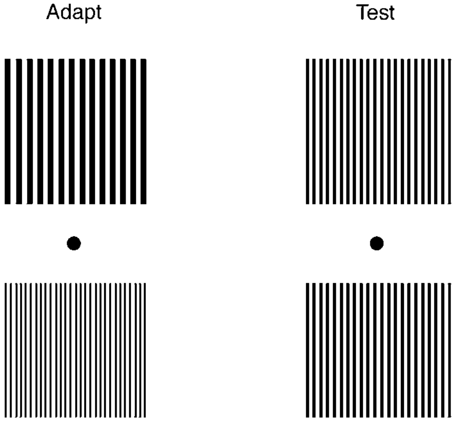


Figure 4.11 The Blakemore–Sutton after-effect. Fixate the left dot for about one minute, then look at the right dot. The upper right grating will then seem more narrowly spaced than the identical lower right grating. (This illusion was first described by Blakemore & Sutton, 1969.)

The model of perception to emerge from the work described in this section is that vision proceeds in two distinct stages. Demonstrations such as that in Figure 4.12 prove that a scene can be physically analysed or decomposed into a set of component spatial frequencies. But it is obvious from Figure 4.12 that this process could be reversed: adding patterns of spatial frequencies would yield complete pictures. And if complete pictures can be formed in this manner, might not the same be true of complete *percepts*? As yet, there is no strong evidence as to where the necessary syntheses take place in perception, neither do we have certain knowledge as to how this process is achieved, although we shall describe some hypotheses concerning this in a later chapter on the computational approach to visual perception. We can say, however, that the researches described above have been a rich source of ideas about vision and that the earlier work on visual analysis is beginning to be complemented by ideas as to how such analyses are later combined to form percepts.

Neural structures mediating responses to spatial frequency

An important question arises over the type of neural structures that might be capable of mediating the responses to spatial frequency described above.

Consider a hypothetical arrangement in which two gratings, one of high spatial frequency, the other of low spatial frequency, were moved across a moderately wide aperture, and a light meter measured the energy reflected from the aperture. What would the meter reveal? Clearly, the stripes of the high-frequency grating are very fine relative to the aperture. Therefore, the meter would simply record the average luminance (or reflectance) from the light and dark stripes. Moving the grating would have no effect upon this average output.

But were the low-frequency grating to be moved across the aperture, it is clear that the larger stripes would exert changes that would be detected. Dark stripes would fill a large portion of the aperture to give a low signal on the meter; light stripes, when they appeared, would produce a sudden change in output. Thus, the aperture is acting as a filter biased to low-frequency gratings. It is easy to see how a smaller aperture would allow the light meter to respond actively to finer (high frequency) gratings. And it is obviously a simple matter to combine different apertures so that they could effectively decompose a complex grating, responding selectively to particular bands of spatial frequencies. Thus, we can consider the aperture as a component in a spatial frequency detection model, with aperture size determining to which frequencies the system is most sensitive.[4]

The question now is whether there is a known neural mechanism that could act in a manner analogous to an aperture. The reader has probably

[4] The author first came across this way of thinking about filters, apertures, and spatial frequencies in Kaufman (1974).

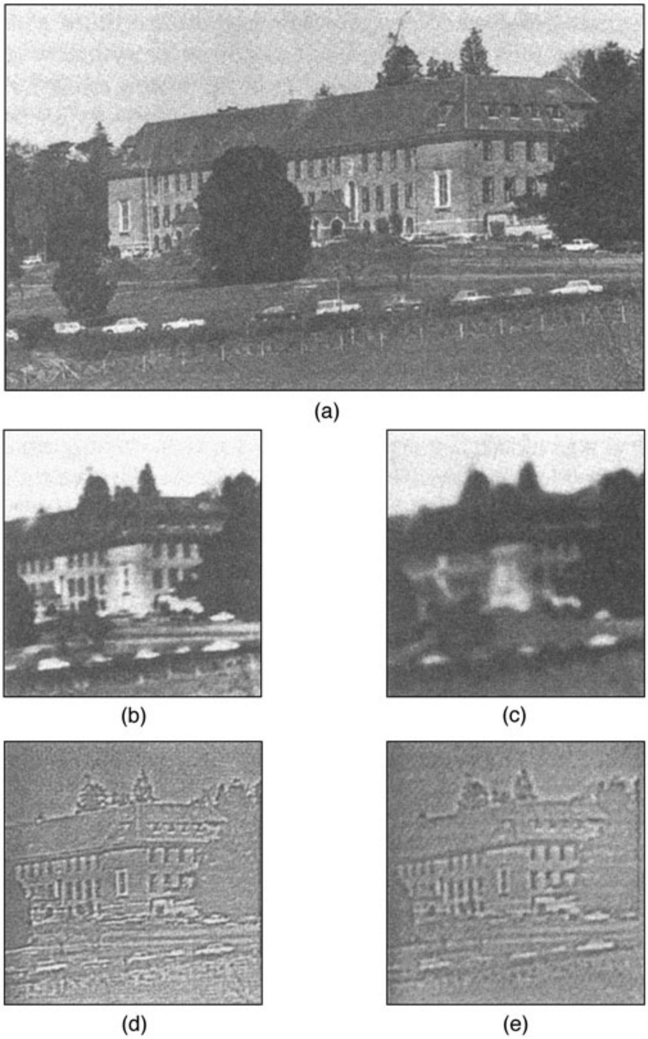

(a)
(b) (c)
(d) (e)

Figure 4.12 Computer-processed images. (a) the original photograph (of a famous psychology department). (b) An area of the original selected for processing, after conversion to pixels. (c) The result of processing the sample through a low-pass filter: the picture now contains only low spatial frequencies. (d) the result of processing (b) with a Laplacian filter. This has revealed the regions in the picture where zero-crossings occur. In this example, the ‘receptive field’ comprised a central excitatory region of one pixel surrounded by eight inhibitory pixels arranged in a square, with weightings of +8 for the excitatory centre and –1 for each of the eight surrounding pixels. (e) is similar to (d) except that the receptive field or mask applied comprised a central excitatory square of 3 × 3 pixels surrounded by 72 inhibitory pixels. (The author expresses his thanks to Professor M.J. Morgan for generously providing the filtered versions for this illustration.)

anticipated the answer to this question: receptive fields could perform this function.

It was for this reason that so much space was devoted earlier to the discovery of receptive fields. They are known to possess many of the properties required for spatial frequency analysis. Remember that those receptive fields discovered to date show various forms of organization: some have on-centre/off-surround arrangements, others the reverse. Receptive fields have different shapes and sizes, and many are known to be orientation-specific. As it can be shown that pairs of slit-like apertures, arranged at right angles to each other, can detect *any* grating (within their frequency range), it follows that we can now begin to understand how the visual system's responsiveness to spatial frequency may be mediated. During scanning movements of the eyes, lines and edges will move across the retina; thus, large numbers of overlapping receptive fields will be stimulated by features within the visual image. There is therefore a very plausible set of mechanisms that could do the required job of spatial frequency analysis, and spatial frequency information can certainly be recombined to form percepts. Perhaps we are starting to learn something very important about the workings of the visual system. This is a very exciting state of affairs.

Of course, no one has claimed that this is a comprehensive theory, even of low-level visual perception. Nothing has been said regarding, for example, the relation between spatial frequency and colour or texture; neither have we referred to depth or motion perception. But there is now a convincing account of how one important aspect of seeing might be mediated, and a major portion of this account uses the language and concepts of neurophysiology.

Classical computer models vs. parallel distributed networks

For the past 30 years or so, the most important model in experimental psychology has been the digital computer. When these machines became widely available in the 1960s and 1970s, many psychologists and others in related disciplines found them irresistible as metaphors for thinking organisms.

Classical computer models

The classical digital computer is often defined in its abstract form as a Von Neuman machine, in honour of one of the pioneers in the area, John Von Neuman. In essence, Von Neuman machines have the following characteristics. There is an important distinction between the permanent structure of the machine, the hardware, and the set of programmed instructions, the software (it has been claimed that this might be analogous to the brain–mind distinction). The machines are rule-governed; they operate via explicit instructions. The operation of the machines is sequential. Knowledge or information within the system is stored in specific memory addresses and retrieved by means of these addresses. Von Neuman machines are controlled by special units, known typically as 'central processing units': to this extent they are essentially 'top-down' systems. Finally, it should be noted that actual digital computers need all their components to be working in order to function correctly: they can suffer catastrophic breakdowns.[5] The listing above is a partial description of both the modern digital computer and the human being, in the opinion of many workers in the modern discipline of artificial intelligence. The achievements of such machines when performing tasks analogous to human mental functions have often been highly impressive. These range from successful object recognition to the playing of chess at Grand Master level.

However, during the years when the digital computer had become the dominant model for psychology, particularly in the related discipline of cognitive science, a few theorists were beginning to have doubts about its suitability as a metaphor of the brain and the mind. The approach described above arose from within artificial intelligence research. Fundamental criticisms of artificial intelligence work have been made by those who doubt whether it is in principle possible for machines to simulate human processes such as thinking and perceiving. Prominent among these critics is the philosopher Herbert Dreyfus.

Dreyfus (1972) and others have marshalled a number of arguments against computer simulations, some of which will be mentioned. They are included here to show that there are those who challenge many of the assumptions underlying artificial intelligence. If these are truly unsound, then the artificial intelligence approach, in its basic form, could be doomed to eventual failure. We shall now summarize some of these doubts.

The first problem arises from a comparison of human performance on the one hand and the workings of neurons on the other. Once it has fired, a neuron goes into a phase when it cannot be made to fire again. This is known as the absolute refractory period of the neuron. There follows a period, the relative refractory period, when the neuron will fire, but only to increased stimulation. The durations of these periods vary, but commonly reported values for the total recovery time of neurons are of the order of tens of milliseconds. Now consider this truth concerning human perception: we can do a great deal in a mere 200 milliseconds. In this small fraction of a second our eyes can make a fixation on a scene or printed page and extract amounts of important information, before starting to move again. Using a tachistoscope, it can be shown that the recognition of faces, words and objects, can all be achieved within 200 milliseconds. We can also make

---

[5] When starting to rewrite this chapter, the author found that his computer was behaving in a bizarre manner. Many sub-routines (macros) had vanished, the computer's date functions had reverted to those of a previous year, and the screen became unreadable. After several hours, a skilled technician traced the fault to a small component which had failed and which was replaced at a cost of only £5.

certain decisions within this period, as is shown by the study of human reaction times: think about how much knowledge and experience is called upon when an airline pilot, suddenly detecting a likely collision, banks the plane steeply to one side.

However, recognition and decision-making are obviously complex activities. In visual recognition, for example, it is necessary to process the visual image and then compare the output with material stored in memory, finally deciding whether or not the item is something one knows. But if such psychological processes comprised strictly sequential stages, as in a digital computer, then the refractory periods of neurons allow us to deduce that there must be a limit to the lengths of these sequences. This has been expressed by workers in artificial intelligence as ‘the hundred-step rule’. It seems highly implausible that the recognition of, say, a familiar face could be carried out by a sequence of only 100 neurons. The obvious implication is that in the brain, in contrast to the digital computer (where speeds of processing are hundreds of thousands of times greater), neurons must commonly act in parallel. This is an important deduction.

Other psychological knowledge, even everyday experience, suggests other ways in which human behaviour differs from the sequential, rule-governed, symbolic manipulation of the digital computer. Humans find it easy to work with ‘fuzzy sets’. For example, we understand what is meant when someone says that a place was ‘crowded’. But how to define ‘crowded’? A crowded telephone booth differs from a crowded restaurant; both differ from a crowded stadium. Humans can see jokes and appreciate puns: it takes a novel combination of different kinds of knowledge to appreciate the wit and power of James Joyce’s ‘Lawn Tennyson’. To ‘see’ a pun like that is to solve a problem. Can one visualize a traditional computer, with its fixed memories and sequential operations, doing such things? Further, this elusive, ‘lateral’ aspect of thought is manifest not only in the appreciation of jokes and puns: Kekulé hit upon the ring structure of the benzene molecule after he had visualized a snake biting its tail; what could be more different than an animal and a molecule, except for this one geometrical similarity?

Humans can generalize. The author’s daughter, then a little girl, once looked up at the rose window in a church and said, ‘It’s a telephone dial’. We see such similarities with ease: clouds can look like faces; penguins like waiters. And when we know a symbol such as the letter ‘A’, we can recognize it in any typeface, at any size, and in any orientation.

Humans can also distinguish between the essential and unessential features of patterns. In solving a problem (and perceiving frequently does require problem solving) the solver must acquire knowledge of what is and is not relevant to the solution. But it is a characteristic of many of the problems that humans can solve that the essentials are not known ahead of time. There are no simple rules for us, nor any for the computer. Where computers have in fact solved problems, it has been the programmers who have stipulated what is and is not relevant.

Humans can take account of context. In Wittgenstein's phrase, 'a mouth smiles only in a face'. The expression is not deducible from a simple list of all the features of the face; the organization of the features *is* the expression. Similarly, humans can use context to complete percepts in ways that are often complicated and sometimes circular. Consider the middle symbol in the words CAT and THE in Figure 4.13. The resolution of the ambiguous central shape must depend in part upon the context supplied by the two familiar words; but the words themselves are formed using the ambiguous symbol. The effortless manner in which we solve this and other perceptual problems seems very mysterious and complex, and not the sort of process to be easily modelled by traditional computer simulations.

Forerunners of the new type of model

During the 1930s and 1940s the distinguished neuropsychologist Karl Lashley set out to discover the sites in the brain where learning and memory occurred. He failed. In a famous monograph entitled, 'In Search of the Engram' (1950), Lashley described a series of experiments on the rat's brain in which experimental damage was caused to different regions before and after learning. The results of these experiments showed that the precise site of an experimental lesion is less important in its effects on the rat's memory or learning ability than the amount of cortical tissue that has been damaged: the greater the damage, the greater the impairment in performance, with complex skills suffering more than simple ones. Lashley elevated these results into the (self-explanatory) principles of equipotentiality and mass action.

Hebb, a contemporary of Lashley's, took a discovery from neurophysiological research and developed it into a possible mechanism for perceptual learning (Hebb, 1949). The discovery was that groups of interconnected neurons continue to show increased activity after the termination of the event that originally disturbed them. Hebb proposed that clusters of neurons displaying this reverberating activity acted as functional units and that modifications to such an interacting network could be the basis of both short-term and long-term learning. Hebb called these networks 'cell assemblies'.

As a final antecedent to the development of connectionist networks, we may cite Rosenblatt's Perceptron (Rosenblatt, 1959). This was a device (in fact, like most devices to be described, it was simulated on a computer) comprising an input layer or 'retina', a decision layer, and between these a

Figure 4.13 Context and ambiguity: 'A' or 'H'?

set of predicates. The retina can be interpreted as a simple array generating binary outputs when stimulated in some manner. The predicate layer comprised a set of threshold units, each connected to a subset of the retinal units and capable of computing some simple function from their outputs. The decision units were joined to the predicate layer by a number of modifiable connections. The task set for the Perceptron was to see to what extent it could adjust its outputs to match a given input: could it act as a primitive pattern recognizer?

The Perceptron was a 'one-layer' computing device (it had only a single modifiable layer). And it contained vital flaws that were subsequently exposed by Minsky and Papert (1969). Nevertheless, it will become apparent that the Perceptron was an important forerunner of subsequent developments in this area.

So far we have not yet said what a connectionist network actually is, other than to sketch the Perceptron as a forerunner. In order to prepare the reader for what is to come, and to give a feel for this type of thinking, we have included Figure 4.14. This is not what is usually meant by a connectionist network. It is included here simply to demonstrate that simple components, connected in certain ways, can do complicated things. In Figure 4.14 the units respond whenever they receive an input. The inputs can be excitatory

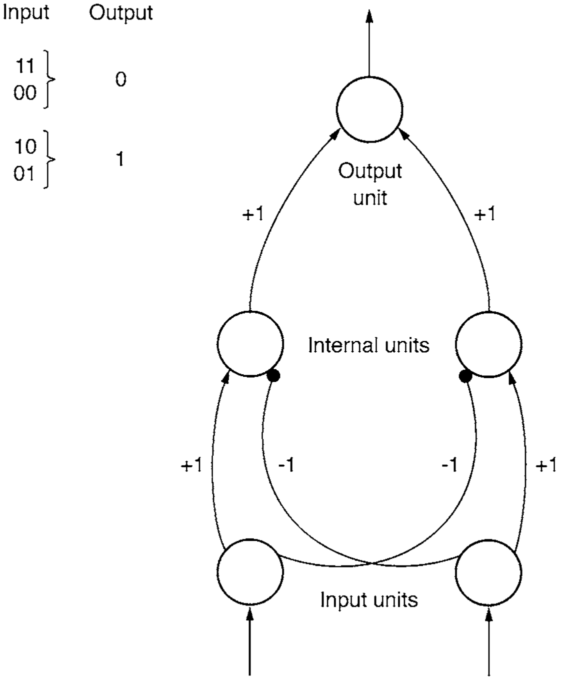


Figure 4.14 Five simple units which, when interconnected by excitatory and inhibitory links, can function as an Exclusive–Or logical gate.

or inhibitory. Consider what happens when the left-hand unit receives an input. It stimulates the left-hand internal unit positively, while inhibiting the right-hand internal unit. As a result, the left-hand internal unit will trigger a response in the output unit. However, a little thought shows that if both input units receive inputs simultaneously, the output unit will not fire. This is because of the mutual inhibition of the internal units. If we now work through the four possible input patterns (left on, right off, and so on) it becomes apparent that the output unit will respond if *either but not both* of the input units fire. What we have achieved with this simple configuration is the important logical function, Exclusive–Or (refer again to Figure 4.2). Now suppose (a) that the units in Figure 4.14 could be given threshold values, enabling them to store inputs from more than one other unit, and (b) that the strengths of the connections between the units could be altered as a result of previous activity. The sense of excitement concerning what such a system might be able to do may be felt already by the reader.

Parallel distributed (connectionist) networks

After that preliminary demonstration, it is time to introduce connectionist networks. Rumelhart, Hinton, and McClelland (in Rumelhart & McClelland, 1986) provide an excellent description of connectionist networks that forms the basis of the following account. Interested readers should certainly consult this invaluable exposition. One word of advice: the terminology in this area varies from writer to writer. The term '*connectionism*' defines the general approach to this new form of modelling. The actions of the models are sometimes described as 'parallel distributed processing'. Some authors use the terms 'neural networks' or 'neural nets'. Beware of these variants when searching the literature. In what follows, we shall use the term 'connectionist network' (or simply 'network' when the context allows).

Rumelhart et al. (op. cit.) state that there are eight important characteristics of any connectionist network:

(1) *A set of processing units*. The units 'represent' some aspect of the real or hypothetical world against which the network will be tested. The representation may be *discrete*, in which case the units will represent particular parts of a display – shapes, letters, or words. Or the representation may be distributed, in which case each unit will represent some small feature-like entity, while the pattern existing among the set of units represents some more abstract aspect of the display or world. Each unit has but one function: to accept inputs from other units, compute some value, and then pass this on to neighbouring units. The computations may occur simultaneously and so this part of the network has a marked degree of parallelism. There are three possible types of unit: input units, such as we have been describing; output units, sending signals from the

system; and 'hidden' units, whose only interactions are with other units within the system.

(2) *The state of activation of the system.* Different kinds of connectionist networks adopt different possible values for the activation of units. These may be discrete, binary (0 or 1), or they may have a range of possible numerical values.

(3) *Outputs from units.* Signals from units to neighbouring units affect the latter. The strength of these effects depends in part upon the activation level of the sending units.

(4) *The pattern of connectivity.* According to the particular model under consideration, one unit may influence its neighbours in different ways, for example, the effect may be additive or subtractive. Or some inputs from units may be given weightings to increase or decrease their influence in the network (note the resemblance between this aspect of the model and a network of interacting neurons). The points to remember are that units interact, and that the strengths of these interactions may be altered by various weighting functions.

(5) *The propagation rule.* In simple terms, this is simply the rule governing interaction or competition between two or more inputs to a unit. If, for example, positive (excitatory) and negative (inhibitory) inputs are allowed to interact in a straightforward manner, then they will cancel when their strengths (or weights) are equal. And so on.

(6) *The current state of the network.* This is an addition to Rule 5. To decide what will result from an interaction between units, we must consider not just the interactions between inputs, as in Rule 5, but how their sum or product interacts with the current state of the unit onto which they impinge.

(7) *Modification by experience.* By this, Rumelhart et al. mean that there are different ways of changing the structure, activity, or 'knowledge' of a connectionist network. Thus, new connections may form and old connections may be lost, or the strengths of connections may be changed through experience. We are now at the core of connectionist network theory.

(8) *Representation of the environment.* The environment in which a model must operate is represented across the input units. Typically, the environment is characterized as a set of labelled probabilities.

As stated earlier, the above description has been abstracted from work by Rumelhart et al., who are pioneers in this field. Their account is much more precise and powerful, but this precision is achieved through the use of mathematical concepts which we have tried to avoid. Nevertheless, readers who ponder over the definitions listed above will achieve an intuitive understanding of these models. To help the reader visualize networks, Figure 4.15 shows an arrangement comprising an input layer of six units, a hidden layer of four units and an output layer of six units. There are more efficient ways

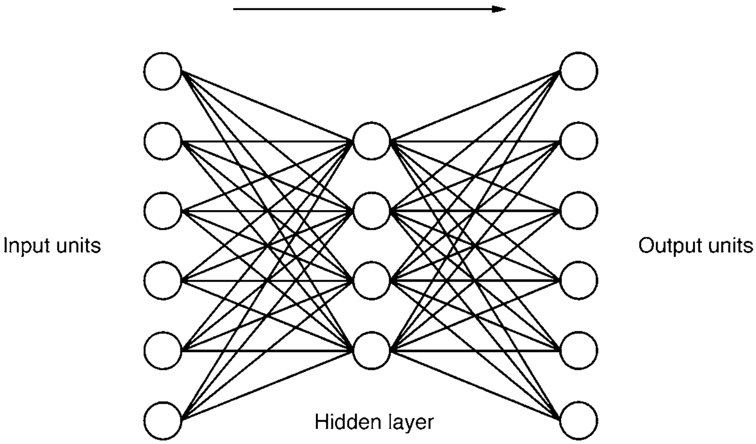


Figure 4.15 A simple neural net comprising six input, six output and four hidden units.

of representing networks that allow the weightings of the various connections to be displayed as clearly as possible, but Figure 4.15 captures the essence of a three-layer network.

Finally, it should be noted that when researchers attempt to model human processes, such as human recognition, their models may comprise two or more networks, each doing the job of one module within the overall model. Thus, one network might function as a (primitive) retina, another might work on the output of this retina, and so on.

In essence, then, a connectionist network comprises sets of simple interconnected units that interact according to weighting rules by which the strengths of their connections can change. The network can be set a problem in the form of an input pattern. A successful solution to the problem comprises a matching or otherwise acceptable output pattern. Obviously, the first 'run' through the network is unlikely to yield the required match or solution; rather, only parts of the output pattern will match the input. How does the network move towards the correct solution? The answer is that the discrepancy between input and output is fed back into the hidden layer. This is done in ways that selectively change the connections to those output units that are correct by strengthening (or 'reinforcing') them according to some predetermined rule. Other connections may be left unchanged or even weakened. In practice, this is a more complicated process than we have outlined, and the assessment of the degree of mismatch between the desired and actual outputs may be in terms of large groups of numbers, grouped as vectors. Those readers possessing the relevant mathematical knowledge can find an explanation of how vector sums and products can be made to operate on a network in the required manner in Rumelhart and McClelland (1986; see the discussion on the use of the delta rule).

There is no supervisory control of the network, no central executive guiding the overall flow of information. And the common distinction between hardware and software vanishes: all units and their connections are essentially 'hard-wired', but the connections can be changed, like software. Further, knowledge is held briefly in the units and for longer terms in the connections: there is no special set of places equivalent to the memory addresses of orthodox computers. Knowledge and memory are not explicit; rather, they are implicit within each pattern of connections. It must be pointed out that very few networks exist as actual machines. Generally, they are simulated on computers. And it is the designer who typically sets the initial weightings in the connections and gives the model its initial problem or input and its target output. The number of trials to solution forms a measure of the performance of the model.

The achievements of connectionist networks

In a later chapter we shall describe the Marr–Poggio account of stereopsis. This very important model was tested by instantiating it in the form of a connectionist network. The model worked.

Other networks have been sufficiently successful to trigger insights of considerable theoretical importance. Networks have successfully solved problems in robotics. They have learned to complete patterns and to recognize faces. They have formed concepts from sets of features and have been able to extract meanings from letter inputs. One network can examine images from a video camera and decide whether or not a railway platform is crowded (remember the earlier remarks on 'fuzzy sets'). The network, named ALVINN ('an autonomous land vehicle in a neural network'; Pomerleau, 1989), was trained by showing it images from a road simulator. The 1200 different images included changes on a large number of parameters, such as road direction, width of road, road curvature, position of obstacles, and many others. After 40 training sessions, ALVINN drove a specially modified Chevrolet van around a university campus.

It would not be wise to describe these achievements as simulating successful behaviour; rather, they represent only pieces of behaviour. Equally, if not more significant, however, are the *emergent properties* that networks have exhibited. Here are three examples.

First, networks have shown *spontaneous generalization*. A network described by McClelland, Rumelhart, and Hinton (1986) was set a classification task. It proved capable of selecting a subset of exemplars, all of which were similar, but none of which had all the formal qualities required for category membership. The network had done something human-like when required to think under conditions of incomplete or imprecise information.

Second, networks exhibit graceful degradation. When given incomplete or slightly faulty inputs, they may not go wildly astray. They may come close to the correct solution to a problem. They get things approximately right. And when subjected to experimental damage (by removing varying proportions of units and interconnections) they do not suddenly fail to work; rather they show a general coarsening of performance, a gradual deterioration which is strikingly similar to the effects of some types of brain damage. In fact, the resemblance between 'damaged' networks and neurological syndromes may be even closer than this. One 'damaged' model actually responded to word probes with words that were unlike the targets in form but similar in meaning, a behaviour which in humans is known as 'deep dyslexia'.

Third, some networks exhibit default assignment. Suppose that a network finds a partial solution to a problem by locating one of the correct target items. The pattern of activity that leads to this solution will raise the strength of certain units and their connections. At the same time, however, the strengths of similar target items will also receive some increments. In this manner, the network can use what it knows to fill in properties of less well-known but similar target items:

> . . . generally speaking, the more similar two things are in respects that we know about, the more likely they are to be similar in respects that we do not, and the [PDP] model implements this heuristic.
>
> (McClelland, Rumelhart, & Hinton, 1986, Chapter 1)

This is, of course, what humans do all the time. To know that a person votes on the right and is in favour of capital punishment gives one a fair idea of his or her views on blood sports. But how interesting that a connectionist network should show similar behaviour as an unexpected, emergent property.

It is of some interest that there are certain characteristics of connectionist networks that may eventually throw some light on the neural systems that inspired their creation. For example, the long controversy over whether perception and other forms of behaviour are innate or acquired might disappear if the connectionist model becomes an accepted model of the brain. The reason is that it is not difficult to see how certain weighting functions in a network could be present at birth, but with the added possibility that they could be modified by later experience. If so, there would still be an interesting set of empirical questions to ask concerning what is in fact present at birth, but the argument between two extreme theoretical positions would be expected to fade away. We have already mentioned this possibility in Chapter 2, when we discussed nativism.

The type of model we have been describing depends upon interactions between simple units. Typically, these models do not have specialized, dedicated centres located within them. Might this throw light on the fact that some areas of the brain outside the major sensory and motor centres have no easily demonstrated localized functions? Damage to the frontal lobes, for example, produces some clear-cut deficits in performance; more commonly, what is observed is a general coarsening of behaviour, an erosion of general ability. This is very different from what happens when an orthodox computer is damaged, but very similar to the behaviour of damaged networks.

We may summarize this short introduction to this new type of model as follows. A network comprising input and output layers joined via one or more 'hidden' layers of simple interconnected units with alterable connection strengths can learn to do many highly interesting things. With their marked degree of parallelism, the networks share some of the characteristics of the brain, and can mimic some of its more interesting properties. The implicit nature of the knowledge within such models – what is represented exists as patterns or relationships between units, rather than as explicitly stored rules – together with the interesting emergent properties of the models in action, suggests that the connectionist network may replace the orthodox computer as a model of the brain or mind. The major insight that designers of networks have confirmed is that *systems using simple components can do very complicated things, provided these components are allowed to compete and interact*. Might this be the way the brain uses its simple components? In one sense, of course, it must be: neurons are what the brain is composed of.

Some more recent work on neural networks

Since the material outlined above was published, research into neural networks has proceeded apace. In fact, there is now a journal, *IEEE Transactions on Neural Networks*, devoted to this topic. Reading recent editions of this journal (which can be formidably technical), one forms the impression that a major thrust in modern research is devoted to improving the performance of networks. A single example must suffice to illustrate this effort. Bartlett, Movellan, and Sejnowski (2002) worked on improving a network's face recognition efficiency. Interestingly, these workers imported some classical statistical techniques into their network model. Pictures of the faces input to the model were in the form of pixels. Now the analysis of such higher-order relationships has long been a tool in searches among variables in the field of psychometric studies of such phenomena such as IQ and possible axes or factors of personality. Combined with the activity of networks, such techniques can enhance performance, as Bartlett et al. have shown. At present, this work has no implications for theories of visual perception; if it continues along the present lines, however, it may be necessary to change our conception of the neural network as a model.

One paper with possible implications for modelling perception is that of O'Brien and Opie (1999). These authors argue that at present the majority of theorists in cognitive science apply computational theory to problems of phenomenal consciousness (computational theory will be explained in a later chapter). O'Brien and Opie argue that instead, modelling should be shifted to a form of connectionism. Thus, degrees of abstractness in conscious experience should be linked to hierarchies of networks in the brain.

These authors go on to say that '. . . phenomenal experience is identical to the brain's explicit representation of information, in the form of stable patterns of activation in neurally realized PDP networks' (op. cit., p.138).

This theory is not yet sufficiently detailed for us to offer any comments upon it, but the fact that it has emerged indicates the amount of thought which is currently being put into the search for possible neurophysiological models of the brain and mind.

Two more recent technical developments

Knowledge of the brain continues to grow at a high rate. And the way this knowledge has grown confirms what was asserted in Chapter 1: namely, that advances in technique commonly lead to new discoveries, which in turn influence contemporary theorizing. In early explorations of the sensory nerves, early anatomists were able only to dissect out the course of, for example, the optic nerve. They found that it went from the eye to the corpus callosum, where there was a complex changeover of fibres; from there the two branches of the nerve went to the lateral geniculate nucleus (a non-cortical structure), and then to the striate cortex at the rear of the human brain. Later, it became possible to cut parts of the sensory nerves, allow for the degeneration of neural tissue, sacrifice the animal and stain sections of the nerve, following the course of the stained track. Knowledge of the sensory pathways thus became more precise. Ablation studies typically removed portions of the visual cortex, allowed the experimental animal to recover, and then noted any deficits in visual performance. Once single-cell recording became possible, it became usual to record the activity of particular visual neurons and then, by increasing the amperage to the fine probe, to kill the particular cell and see what, if anything had changed in the animal's perception.

The discovery of the electro-encephalograph early in the twentieth century made it possible to record potentials from the skull of living humans without intruding in any way into their brains; this technique led to important advances in the understanding of sleep, consciousness, and several related functions. As this work was developing, some surgeons, such as Penfield (Penfield & Rasmussen, 1950), were developing techniques whereby conscious patients, receiving local anaesthetics during brain operations, were able to report their subjective experiences as Penfield stimulated parts of the cortex with electrical probes.

In recent years, techniques for exploring the conscious brain without any risk of damage have been developed. It must be repeated here that the motivation for developing these techniques (and their research funding) arose for medical reasons. Understandably, the search has been to discover what happens to the brain following strokes, the course of Parkinson's disease, multiple sclerosis, and similar disorders. And this is how it should be. However, workers in perception have been increasingly able to join in this research.

A number of new techniques have arrived on the scene in recent years. We shall describe just two of them. The reader will quickly see their significance as tools for understanding brain function.

Position emission tomography (PET)

PET involves the use of tracer substances, such as 2-^{18}F-deoxy-D-glucose. When injected, this chemical moves into cells in a similar manner to normal glucose, but it cannot be metabolized after its initial entry and is as a result trapped in the cells. A radioisotope such as 2-^{18}F-deoxy-D-glucose has a very short life. It emits a positron (a positively charged electron) as it decays. When this positron collides with an electron, the two particles annihilate each other and produce two photons. These photons can be recorded by detectors placed around the patient's body. It then becomes possible to construct a picture of the organ into which the isotope has been injected.

Functional magnetic resonance imaging (fMRI)

fMRI is more difficult to describe. Although the basic principles have been known for many years, it is only relatively recently that it has come into common use in medical diagnosis and neurological and psychological research.

Atomic nuclei spin on their axes. Because the spinning causes them to act as gyroscopes, the nuclei precess (tilt) about their axes. As the spinning nuclei have positive charges, they also act as tiny magnets, with their north and south poles aligned along their spin axes. If now a strong external magnetic field is imposed on the nuclei, they line up with the north poles pointing in a 'southwards' direction. Then, when a radio frequency pulse is applied at right angles to the spin axis of the aligned nuclei, some of the nuclei will tilt away from this axis. The nuclei that tilt are those that resonate with the frequency of the radio input. At the end of the radio frequency pulse, the affected nuclei gradually become aligned with the general spin axis again. This process is known as the 'relaxation time'. To the non-specialist, the next step in this process is the really surprising and interesting one. As the nuclei relax, each acts as a miniature radio transmitter. As nuclei in different tissues transmit different radio signals (depending upon the local environment) their miniature signals can be used to form *images*. A sequence of such images constitutes a series of slices through living tissues.

There are of course many elaborate variations in the techniques outlined above, and we have omitted any mention of, for example, Fourier analysis, which is commonly employed in a number of procedures. The characteristic they share, however, is that they enable pictures to be built up that show the detailed anatomy and functioning of various parts of the body, including the central nervous system. So here, for the first time, some psychological researchers have access to non-invasive, non-harmful, precise techniques with which to study the functioning of the central nervous system. And this research can be carried out while the person whose nervous system is being examined is conscious, capable of carrying out instructions and performing mental tasks.

Here now are few examples of the use of the above techniques to examine a theory.

The scientist Edwin Land (1909–1991) published a series of accounts in which he described a new approach to the explanation of colour vision phenomena (see e.g., Land, 1985). The easiest way to gain an insight into Land's thinking is to carry out a very simple demonstration, using oneself as observer. The reader should obtain two photographic filters, one red, the other blue – in other words, filters selecting the long and short wavelength ends of the spectrum, respectively. With the red filter over the eye, the reader should now look at a familiar scene. The results may be surprising: red objects in the scene will appear very light; blue objects will appear very dark. The reverse will be the case when the blue filter is placed over the eye: blue objects will look light, red objects dark.

Findings such as these encouraged Land to develop what he called the Retinex theory of colour vision. In essence, the theory proposes that colour is determined by lightness values obtained via three filters – in this case, the three cone pigments in the retina. To support this claim, Land created a series of displays, each comprising a set of coloured rectangles. He named these 'Mondrians', after the Dutch artist of that name. Land then illuminated one of his Mondrians with long-wavelength light and scanned the surface with an instrument that measured the lightness of each coloured patch. He then repeated the measurement, illuminating the Mondrian with middle-wavelength and then short-wavelength lights.

Using a special algorithm, Land was able to give each patch of the algorithm a value, based on its reflectance value relative to other patches in the Mondrian. This was repeated with the other two illuminants. As a result of this measurement exercise, Land was able to assemble a three-dimensional model which correctly predicted the appearance of each patch in the Mondrian when illuminated with equal mixtures of his three coloured lights, i.e., with white light.

Land then went a step further. He picked out one rectangle in the Mondrian as it was lit by balanced amounts of light from his three projectors. Let us assume that this patch looked yellow. He then measured the return spectrum of a red patch in the display under balanced light and found, unsurprisingly, that the reflectance values differed markedly from those obtained from the yellow. The final step was to alter the outputs from his projectors, so that the original yellow now reflected the same profile as the red rectangle. The result? The yellow rectangle remained yellow in appearance, and the red, now viewed under very different conditions of illumination, remained red (see Land, 1985, for an account of his work).

Land came to London and showed his demonstration to the distinguished scientist Semir Zeki, who was intrigued. Zeki and his colleagues, using the new techniques outlined above, together with data from psychophysical experiments, appear to have located the place in the visual cortex where the computations proposed by Retinex theory might be carried out: this is V4, a region anterior to the primary visual areas in the brain. Further, they have demonstrated the existence of a cell that apparently shows colour constancy: that is to say, the cell does not respond to the spectral characteristics of a small region of light arriving at a point on the retina, but is influenced by light stimulating surrounding regions of the retina – as Land predicted. Further details of these developments can be found in Zeki (1993).

A very recent piece of research by Kan, Barsalou, Solomon, Minor, and Thompson-Schill (2003) has found equally interesting implications for general psychological theory. These workers employed fMRI techniques to discover which part(s) of the brain are most active in tasks involving mental imagery. The volunteers in this study carried out a series of property verification tasks as their brains were being scanned. For example, when asked questions such as, 'is "frosting" part of "cake"?', or 'is "cow" part of "vehicle"?', volunteers had to press one of two buttons representing yes or no answers. Reaction times were measured.

The interesting finding to emerge from this research was that the most active part of the brain during such problem-solving was the left occipitotemporal region, an area known to be involved in visual object recognition and visual imagery. Kan et al. offer the following hypothesis: *semantic knowledge may be grounded in perceptual regions of the brain*. This is a very challenging idea and one with important implications for psychological theories.

What exciting prospects these new techniques offer!

Some problems with the neurophysiological approach to perception

Description vs. explanation

Descriptions of the behaviour of cells in the visual system do not *explain* that behaviour. Why, for example, is it adaptive to have cells in the visual cortex responding selectively to images of hands on the retina? Frisby (1979) has stated that:

> . . . it is dangerous to assume that a property of a neuron can be directly equated with its functions; 'line detectors' have the property of responding optimally to a line of a given type but this does not mean that they are 'line detectors' in the full and proper sense of this term, that is that their function is to detect lines.

This problem is partly to do with levels of description, a topic that will be dealt with at length in Chapter 8, when Marr's computational theory will be described.

Neurophysiological discoveries and subjective experience

Here is a problem arising from research on a sensory modality that is probably much simpler than vision, namely pain. Much is now known about the neural mechanisms involved in pain. There is a powerful theory – the gate control theory (see, e.g., Melzack & Wall, 1965) – that explains many of the facts about pain and that can be used to guide treatment. However, the thing about pain is that it *hurts*. And pain perception is influenced by many factors, including context: a playful slap can be received with glee; the same slap in a context of anger may arouse strong negative emotions.

In a similar manner, the basic mechanisms underlying the perception of colour are steadily being unravelled. But colour is, among other things, an experience, and one which most of us enjoy. However, since the earliest days of scientific work on the nervous system it was obvious that no actual colours are carried through the nervous system; rather, colour information is transmitted by neural codes. (By analogy, the colours seen on a television screen are generated by pixels capable of generating mixtures of three primary colours. But the signals to these pixels are in the form of numbers: this is how colour is processed by the inner workings of the television set.)

It may of course be the case that phenomenology and the workings of the neural system will never come together. Further, there are mental processes that cannot be detected in conscious experience. Consider being asked to multiply 14 by three. The present author does this by multiplying four by three, remembering the number two and carrying one over when multiplying the final one by three. It is not hard to follow this process in one's head. In contrast, riding a bicycle involves calculations that are mathematically far more complicated than simple multiplication. And one has no inkling of these calculations when riding the bicycle. If the brain could be scanned during bicycle riding (it would be a bizarre experiment), what would the results tell us – apart from the fact that certain areas of the cortex and cerebellum were highly active? Thus, there would be a serious question of the interpretation of the fMRI data.

Before we abandon this topic it is worth repeating that while most of this chapter has been about the *analysis* of perceptual inputs, work on spatial frequency perception has begun to find that neural processes are capable of synthesis. Furthermore, experiments described earlier by Zeki and colleagues have found a cortical cell that exhibits colour constancy; colour constancy is an important part of our subjective experience. These facts may provide grounds for at least some cautious optimism about the relationship between how we see and what we see.

Mind and brain

A portion of this chapter has been devoted to connectionist networks. There are two reasons for this. First, the networks are plausible models of the activity of actual neural mechanisms. Second, the networks have shown themselves capable of mimicking important aspects of human performance. But are they models of the mind, or solely of the brain?

Many, if not the majority, of workers in cognitive science believe that many human functions are rule-governed. This appears to be particularly true in the case of much human reasoning, in language acquisition and use, in some areas of perception, and in human development generally. To date, many of the most powerful explanations of general human performance have adopted the von Neuman formal model of the classical computer. As was described earlier in this chapter, the von Neuman computer follows rules, works in stages that are organized in a hierarchical manner, and operates according to routines and sub-routines. Explanations of, say, the visual or auditory perception of words commonly use flow-charts showing how the various stages interact to produce the final outputs. These are classical, rule-governed models.

So we have a dilemma here: if the brain is organized according to the network models, which are not rule-dominated, how is the mind organized?

One distinguished philosopher, Daniel Dennett (Dennett, 1991), has faced up to this problem. Knowing about research in cognitive psychology, but aware of the functions ascribed to neural networks, Dennett has come up with a challenging idea. Might it be the case that cortical tissues, behaving in a manner suggested by connectionist network theory, create a *virtual* von Neuman machine? This is a remarkable and highly stimulating attempt to resolve the major issues discussed above: the brain works as best it can, given the properties of groups of neurons, but in doing so, supports the existence of a virtual rule-governed machine with all the potential that such a machine offers. Dennett's hypothesis concerns consciousness generally, but it can obviously be used in thinking about perception:

> Human consciousness is *itself* a huge complex of memes[6] (or more exactly, meme-effects in brains) that can best be understood as the operation of a '*von Neumanesque*' virtual machine *implemented* in the *parallel architecture* of a brain that was not designed for any such activites. The powers of this *virtual machine* vastly enhance the underlying powers of the organic *hardware* on which it runs . . .
>
> (Dennett, 1991, p. 210, emphases in original)

We end this chapter by asserting, once more, that reading of the discoveries

---

[6] The term 'memes' comes from Richard Dawkins's work on the cultural evolution of ideas, in contrast to the evolution of mechanisms and species (Dawkins, 1989).

that form the material of the present chapter is an exciting experience. These major researches will have a permanent place in the history of perception. And even if it becomes accepted that the true explanation of perceptual phenomena cannot be arrived at using only the language of neurophysiology, then this too may be an important step in our understanding.

Endnotes

- There are many textbooks on the structure and function of the nervous system. Rosenzweig, Leiman, and Breedlove (1999) is a very good example and comes with an illustrative CD suitable for Windows or Macintosh computers.

- As a starting point in understanding parallel distributed (connectionist) networks, readers should consult Rumelhart and McClelland (1986).

- Connectionist networks are being used to develop models of infant development. The problem here is unlike the task of, say, teaching a network to drive a vehicle, where the input and output of the network can be defined; where would a 'target' for a learning infant come from? The way around this problem is to make the target identical with the input.

- Bruce, Green, and Georgeson (1996) contains a superb account of how a parallel distributed network is able to solve a problem in stereoscopic vision.

- Those readers who do not have access to specialist libraries should consider using the Web if they wish to learn more about recent research in vision and the development of new techniques, such as fMRI scanning. For example, Zeki, who featured in the later parts of this chapter, has his own Web page. The present author's favourite search engine is Google, from which he has found a lot of material concerning neurophysiological research. This site was equally valuable in preparing the material on information theory and Kolmogorov complexity outlined in Chapter 2.

- Part of the impressive range of techniques available to modern workers in neurophysiology is displayed in Blakemore (1990).

- Edwin Land, whose work on colour vision has been outlined in this chapter, left college before he was 19. A short time later he had invented Polaroid and founded the Polaroid Corporation. In the late 1940s Land invented the Polaroid camera. In his lifetime Land was granted 500 US patents – second in number only to those granted to Thomas Edison.

Land was eventually awarded the Medal of Freedom, the highest civilian honour in the USA.

- In writing about some of the technical topics in this chapter, we may have over-simplified the attitudes of some researchers. Certainly there are those who prefer to search for neural mechanisms as explanations of perceptual phenomena, rather than to speculate about psychological hypotheses. But the men and women engaged in neurological research are highly intelligent and are well able to discuss alternative approaches to perceptual problems.

- The debate over three-factor vs. opponent process explanations of colour vision is described in most general perception textbooks, including Sekuler and Blake (1985).

- Sekuler and Blake should also be consulted for their excellent descriptions of receptive field and spatial frequency research. Kaufman (1974) contains a very clear and interesting exposition of the relationship between spatial frequencies and the filtering characteristics of various sized apertures.

- Readers who enjoy polemical writing should read the criticisms by Dreyfus (1972), concerning the idea that the computer can serve as a model of the mind. Some of these criticisms have been incorporated into the present chapter.

- Readers wishing to learn more about parallel distributed processing or connectionist networks should first read Orchard and Phillips (1991), which is a beginner's guide. The guide is supported by computer software which makes it possible to gain hands-on experience with various connectionist networks.

- McClelland, Rumelhart, and Hinton (1986) is an advanced, comprehensive and authoritative exposition of the general topic of neural networks.

- The power of modern neurophysiological research on vision is evident throughout Blakemore (1990). Beware, though: some of the technical work described is fiendishly complicated.

- Reference is made in this chapter to Julesz's concept of the texton. Professor Richard Gregory has pointed out to the author that Julesz eventually became disappointed with his theoretical creation. It is certainly true to say that he did not develop his original ideas very much further.

5 Empiricism: perception as a constructive process

> . . . whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our head.
>
> (William James, 1890)

In 1959 a Canadian psychologist, Robert Sommer, described an incident leading to a trial at which he was a juror. Here is Sommer's account of the incident:

> A hunting party went out one afternoon looking for deer. While driving through a field, their car became stuck in the snow and eventually the transmission broke. Of the five men in the party, two volunteered to go to a nearby farmhouse for help. Of the remaining three, one remained in the rear seat while the other two stood at the front of the car. Meanwhile, one of the two men on the way to the farmhouse decided that there was no reason for both of them to go, and he thought he might be able to scare up a deer. Unknown to the men in the car, he circled around down a hill in front of them. At that point, one of the men standing outside the car said to the other, 'That's a deer, isn't it?', to which the other replied in the affirmative. The first then took a shot at the deer. The deer pitched forward and uttered a cry, which both men heard as the cry of a wounded deer. When the deer started running again, the second man implored, 'Don't let him get away, please get him for me'. The first man fired again and the deer went down but continued its forward movement. A third shot brought the deer to the ground and both men started running towards it. By this time the third man in the car, who had been trying to find and focus his field glasses, suddenly called out, 'It's a man'.
>
> (Sommer, 1959)

The ideas in this chapter attempt to explain what happened on that tragic afternoon.

After that real-life story, consider two simple demonstrations. Look first

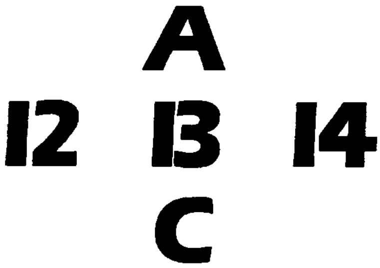


Figure 5.1 Ambiguous stimuli: '13' or 'B'?

at Figure 5.1. What do you see? Most readers will interpret the figure as a row of numbers and a column of letters. Now notice that the numeral '13' and the letter 'B' are in fact identical. Interpretation of the patterns is affected by their context.

Turn now to Figure 5.2. For many readers this will appear as a random jumble of black shapes on a white ground. If this is what you see, your perception of the figure will now change.

The first hint is that the figure contains a face: can you see it? If not, then note that it is a Christ-like or Cavalier face occupying the top third of the rectangle and looking out of the page. Now can you see it? If not, note that the figure is strongly lit from one side, has two penetrating eyes, long hair and a beard. If you still cannot see the figure, look at Figure 5.3, which contains a more explicit plan of the face. If you happened to see the figure

Figure 5.2 The hidden face. See the text for hints on how to see the face. (From Porter, 1954, *American Journal of Psychology*. Copyright © 1954 by the Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.)

immediately, it is worth showing it to a friend and then helping him or her to see it by a series of hints. Seeing the face can occur quite dramatically and is a fascinating experience. Once seen, the face will always emerge from Figure 5.2.

It is clear that in both these demonstrations one's perceptions are not predictable simply from the parts that form the 'B', the '13', or the face. Something – the context, the hints – seems to have come between the registration of the stimuli and our final response to them. Is this an essential part of perceiving?

Many who have worked in the field of perception, possibly the majority, have not been committed solely to one particular theory or approach. Researchers have had their imaginations triggered in a variety of ways. They may have read about some new phenomenon and decided to set up equipment to enable them to see it for themselves. It is then a small step to make small changes and explore their effects. Soon, a research programme for the next few years has crystallized. Similarly, people can simply notice something odd about their own perceiving under unusual or unfamiliar conditions and decide to investigate what nature has tossed into their laps. And much research is still of the 'What if . . .?' variety. What if chimpanzees could learn sign language? What if observers became weightless? What if an animal is prevented from using its eyes for the first three weeks of life? And so on.

In this sense it is almost possible to be atheoretical in perceptual research. Almost, but not quite. The framework within which one thinks, the attitudes implied by particular experimental designs, and even the ways in which one expects observers to respond, are all subtly influenced by the current Zeitgeist. All who study perceptual phenomena have some beliefs concerning the fundamental nature of perceiving. To this extent, nobody is really atheoretical. When such a set of beliefs and assumptions becomes widespread and strongly influential, the term 'paradigm' may be invoked. The impact of Darwinism on biology is a clear example of a paradigm shift, as is the more recent impact of quantum mechanics on theorizing in physics (see Kuhn, 1970, for a discussion of paradigms in scientific research).

We now assert that the dominant paradigm in perceptual research in the twentieth century has been empiricism, which is the subject of this chapter. The thoughts that may have been suggested by the opening demonstrations at the start of this chapter are the same as those which led to the spread of empiricism in psychology: the idea that perception is something more than the direct registration of sensations; that somehow other events intervene between stimulation and experience. We shall attempt to show just how fruitful this idea has been.

The approach known as empiricism is based on doubts as to whether proximal stimuli can adequately represent distal stimuli; rather, it maintains that the brain has to do much in order to gain true knowledge of the world. At this point it is important to warn philosophically informed readers that

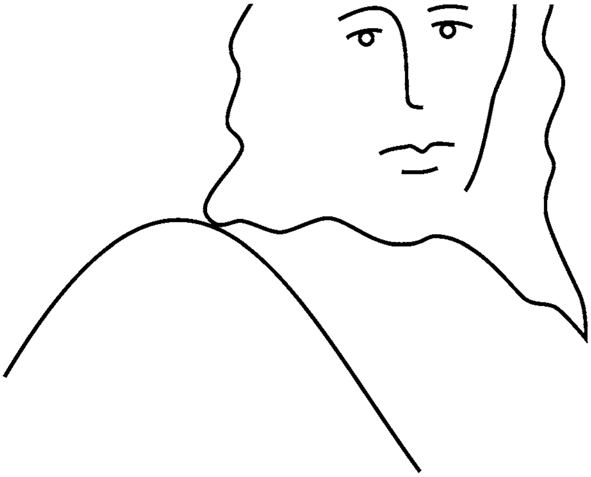


Figure 5.3 A key to the hidden face.

empiricism in psychology is *not* identical with the tradition which developed in the writings of Locke, Berkeley, Hume, and Mill – the British empiricists. Locke, for example, was concerned with the origins of *ideas*. His writings stress that ideas can come into the mind only as the result of experience. Locke did not argue, as is sometimes claimed by psychologists, that we have to learn to see. But, as we shall show, some psychologists have claimed that this is indeed the case. Empiricism in psychology is therefore somewhat coarser and more extreme than the philosophical version. An excellent discussion of the topic will be found in Morgan (1977).

The remainder of this chapter will cover the following topics:

- The historical background to empiricism.
- A modern version of empiricism: Gregory's theory that perceptions are hypotheses.
- An evaluation of Gregory's theory.
- Other modern versions of empiricism.

Historical background

Helmholtz

Hermann Helmholtz (1821–1894) became a doctor of medicine by the age of 21. In 1847 he published a paper entitled, 'Über die Erhaltung der Kraft' ('On the conservation of energy'), which altered the direction of physics for decades to come and was the basis of the new science of thermodynamics. He went on to hold various prestigious academic posts, including the Chair of Physiology at Heidelberg, and the Chair of Physics in Berlin. He published more than 200 papers on mathematics, medicine, anatomy, physics, philosophy, physiology, and psychology.

Helmholtz was one of the founders of perceptual research, and was probably the most gifted, original and successful worker in perception to date. The list of his discoveries and inventions is staggering: the first scientific account of hearing; the first scientific explanation of musical effects; probably the best book yet on seeing, *The Physiological Optics* (1909–1911); a major theory of the workings of the inner ear; a major theory of colour vision; the invention of the ophthalmoscope – Helmholtz was the first person to look into a living eye.

Helmholtz and unconscious inference

In mid-career, Helmholtz discovered an interesting problem. He had noticed that if a small piece of grey paper is laid over a red surface, it becomes tinged with green. Where does this green come from? Subsequent discoveries permit an explanation of the induced green in terms of known physiological mechanisms. But Helmholtz knew only that green is the *complement* of red (in colour vision research, two hues are said to be complementary if their light when mixed yields an achromatic grey). Helmholtz therefore offered an explanation of the induced green in terms of the viewer's knowledge of the fact that the grey paper, when adjacent to the contrasting red, should yield its complement.

With this explanation of a perceptual effect, Helmholtz brought empiricism into experimental psychology. He argued that between sensations (when our senses first register the effects of stimulation) and our conscious perception of the real world, there must be intermediate processes of a constructive nature. These processes resemble thinking, in particular inferential thinking, and because of them perception can go beyond the evidence of the senses – evidence which is often inadequate or distorted. Put another way, if distal and proximal stimulation are not identical, then intermediate processes must exist: how else to explain the veridicality of perception? But when we introspect, we are not normally aware that we are making inferences, neither can we change our perceptions at will. Therefore, the inferential processes must be unconscious.

Armed with this simple idea, one can begin to explain a variety of important phenomena. For example, if the brain can calculate object distance, possibly by a process resembling triangulation, then this might be a way of compensating for the reduction of retinal image size with distance. This would provide a basis for size constancy.[1] In a similar manner, it is possible

---

[1] Size constancy is interesting when it breaks down. The author's son, when a child, once asked how people could fit into a 'tiny' aircraft flying overhead. Over a hundred years earlier, Helmholtz had recounted his childhood puzzlement over the doll-like appearance of people high above him in a belfry.

to understand why rectangular objects maintain their apparent shape when viewed obliquely. And so on.

Helmholtz did not fully explore the implications of unconscious inference. He offered some rules about the inferences – that they were the result of associations and experience, that they were inferential, that they were the result of association and experience – but the central idea is not fully developed in his subsequent writings. However, the idea of unconscious inference did enable Helmholtz to avoid nativist solutions to the problems of perception (we see things at their proper size because we are built to see in this way, which is hardly an explanation), and this may have been his main aim.

Helmholtz's prestige, added to the basic appeal of empiricism, made this way of thinking about perception almost irresistible for many workers in perception. The acceptance of his ideas was also reinforced by the 'inference revolution' described in Chapter 3. There was in fact some opposition to Helmholtz's views during his lifetime. For example, another great physiologist, Hering (Helmholtz's contemporary and rival), adopted a nativism that was totally opposed to empiricism; and, as we have seen, the Gestalt movement had little sympathy for the idea that perception was based upon associations. But despite these important exceptions, a major paradigm had emerged: perception was to be thought of as an indirect, constructive, inferential process.

Empiricism after Helmholtz

We will now describe a selection of some of the work during the years between Helmholtz and the present that has reinforced the empiricist conception of perception.

Attention and set

In a classic study of human attention, Külpe (1904) used a tachistoscope to deliver brief exposures of displays of variously coloured letters. The observers in the experiment were directed to attend to some aspect of the display, say the position of certain letters. When asked subsequently to describe some other aspect of the display, for example the colours of the letters, they were unable to do so. The significance of this famous demonstration is that although all the information from the brief display must have reached the eye, at some point between the formation of the retinal image and the production of the final report, selection had taken place: what is taken in from a display depends not only upon the properties of that display but on the 'set' the viewer has adopted. Perceptions are not simply inputs.

Drives and perception

Sanford (1936) showed ambiguous pictures to groups of school children and asked them to write down what they had seen. The experiment was run at different times of day. It was found that twice as many food-related responses were made before compared with after meal times. Hunger can influence what is seen.

The influence of stereotypes

Early in his famous book *Remembering* (1932), F. C. Bartlett describes a demonstration that he ran during the first open day at the new psychology laboratory in Cambridge. Visitors were asked to look into a tachistoscope and report what they could see. The picture in the tachistoscope was that of a man wearing a naval officer's cap. Many viewers reported the presence of the cap but added, wrongly, that the man had a beard: the current stereotype of a British naval officer. Prior expectations had influenced what they had seen. (In the subsequent researches for which he became famous, Bartlett was able to show that stereotypes and expectations exert an equally striking influence on long-term remembering.)

The New Look experiments

In the years following the Second World War, a group of American psychologists, many of whom shared an interest in Freudian or other psychodynamic theories but had also received training as experimentalists, reported a series of researches, which became known, collectively, as the New Look psychology (after a popular contemporary fashion in clothes).

Bruner and Goodman (1947) carried out an investigation into children's ability to judge the size of coins and found that the perception of size was influenced by the value of a coin, the effect being greater with children from poorer homes. It was then reported that it took longer to recognize 'taboo' words than control words when these were presented tachistoscopically (McGinnies, 1949). More dramatically, Lazarus and McCleary (1951) reported that after certain words had been paired with electric shock, they took longer to recognize, and that even before their recognition thresholds were reached, the words induced physiological responses in observers.

The New Look experiments are now generally discredited. For example, might not the delayed responses to taboo words be a response effect: the observer wishing to make sure of being correct before uttering the taboo word? And poor children would be expected to be less familiar with high-value coins. However, publications like those referred to could only reinforce the idea that perception is a constructive process: they are part of the story.

The Ames demonstrations

During the 1940s, Adelbert Ames of the Dartmouth Eye Clinic, Connecticut, developed some of the most compelling illusions ever seen. The most famous of these are the Ames room and the Ames window (see Ames, 1949).

The Ames room is shown in plan view in Figure 5.4. It is an irregular shape with a receding rear wall. But the room is decorated in a special manner: the patterning is such that it projects an image to a viewing point in the front wall that is identical to that which would be produced by a wall at right angles to the two side walls; in other words by a normal wall.

When one looks into the Ames room, one sees it as normally proportioned. But if a person walks from one of the two far corners of the room to the other, the room stays rectangular *but the person appears to change size*. This wonderful illusion does not disappear when one learns the true shape of the room.

The Ames window is simply a trapezoidal shape with a window design added to it in such a way that when the window is viewed obliquely from 45° the outline and the details appear rectangular. When such a window is

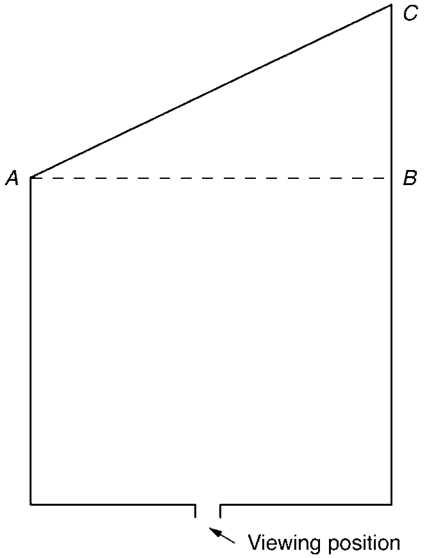


Figure 5.4 The Ames room in plan view. The wall *AC* is shaped and decorated so as to be in position *AB* when seen from the viewing position. Viewed monocularly from the front, the room appears rectangular. However, an object moving from *A* to *C* will appear to shrink.

rotated one sees, not rotations, but *oscillations*, the window appearing to change its direction of rotation half way through each cycle.[2]

The original explanations of these two Ames illusions were in terms of our familiarity with rectangular rooms and windows. It must be said that this explanation has been challenged: see, for example, Day and Power (1965). But note how powerful a demonstration of the vulnerability of perception is contained in these ingenious inventions. How could one deny that perceptions are modifiable constructions, rather than direct responses to patterns of stimulation, when one has been made to doubt the evidence of one's senses in such a captivating and compelling manner?

Ames believed that we carry around with us knowledge of the 'typical' size and shape of familiar objects, which we usually see over a very restricted set of distances. If unusually large or small versions of familiar objects are illuminated in the dark, their apparent distance is determined partly by their actual distance, partly by the visual angle they subtend under natural conditions. Similar explanations can be advanced to explain illusions such as the Ames room and the Ames window. Thus, perception is a dynamic interplay between current stimulation and expectations based upon our previous dealings with the world.

Day's work on the geometric illusions

Many years ago, Bruner and Postman published an account of an interesting perceptual experiment in which playing cards were exposed to observers for very brief periods of time. Each observer's task was to identify either the colour of each card, or its suit. Unknown to the observers, the sequence of stimuli included a few trick cards, in which the colours had been changed: for example, the King of Spades might appear as a red card. Not surprisingly, perhaps, it took the observers much longer to identify these trick cards correctly. However, in some cases when, say, the ambiguous King of Spades was presented, observers reported that the colour seen was 'brownish'. This was obviously some sort of compromise between what the observers knew about playing cards and what they had actually seen.

The distinguished Australian psychologist, Ross Day,[3] has subsequently developed a theory of perception that accounts well for the type of finding described above.

Although we cannot prove this point, we will hazard a guess that the majority of workers in visual perception share this common belief: when we know how visual perception achieves a high degree of veridicality under

---

[2] The present author built an Ames window and showed it to many students. He still experiences the illusion.

[3] The present author has had many profitable discussions about the nature of perception with Professor Day over the past 20 years.

normal conditions, then the errors into which perception can be lured by visual illusions will be completely explicable. This was certainly the position adopted by American psychologist J. J. Gibson, whose work is the subject of Chapter 6. Day is sympathetic to this view, but argues as follows:

> While not dissenting here from that view, it can be suggested that the opposite standpoint is equally sustainable – that if a satisfactory explanation can be found for perceptual illusions an explanation of veridical perception will follow
>
> (Day, 1989).

A large portion of Day's career has been spent analysing and manipulating a group of figures known as the 'geometric illusions'. One of the most famous of these is the Müller–Lyer illusion, which was shown in Chapter 2 (see Figure 2.7).

Day's approach is essentially cue-based. That is to say, our way of accessing the external world (and self-induced activity) is via the spatio-temporal patterns of energy generated at our receptors by features and events in the external world. Perception may depend upon a single cue, or it may be associated with many. But cues may also be 'contrived': they can operate in the absence of those physical features with which they are normally correlated. Thus, the basis of stereoscopic vision is the disparity between right- and left-eye views of the world, which comes about because our eyes are in different lateral positions in our heads. However, when in 1838 Wheatstone invented the first stereoscope, he simply drew pictures containing disparity information. The result was that, when fused in his stereoscope, Wheatstone's flat pictures generated depth. However, depth seen in a stereoscope display is never as good (clear, strong, vivid) as that seen in normal vision. The obvious presence of the flat paper on which the stereograms are printed reduces the sense of depth. The result is a perceptual *compromise*. To see how Day explains various geometric illusions using his compromise theory, see, for example, the following papers: Day and Kasperczyk (1985), Day and Duffy (1988), Day (1989).

Attention and perception

Some very famous studies of human attention were published during the 1950s and 1960s. We shall not attempt to review this large literature. It must be stated, however, that influential books by Broadbent (1958) and Neisser (1967) made a powerful case for the selective nature of much human perceiving. An observer, asked to monitor one of two aural messages delivered simultaneously, one to each ear, will subsequently be unable to say very much about the other. The observer will, however, hear his or her own name in the non-attended ear. In another widely used experimental situation, an observer who has just scanned rapidly through a visual array for a target will be able to say very little about the non-target items scanned through. These and many other reliable effects illustrate the selective nature of perceiving, and seem to show that perception can come under the control of central factors and is not determined solely by local conditions of stimulation. This is very suggestive evidence from the point of view of those who support an empiricist view of perception.

This, then, was the state of thinking concerning the indirect, constructive nature of perception 100 years after Helmholtz had formulated the doctrine of unconscious inference. To summarize, empiricism in psychology conceives of the perceiver as being not unlike the captain of a submarine. He has knowledge of the medium in which he is submerged but cannot experience it directly. So it is necessary to plan according to the knowledge that experience and training have provided. From time to time indirect samples of the environment are taken: instruments show the distance from the ocean floor, the vessel's heading, the presence in the area of other submerged objects. The better (the more alert and experienced) the captain, the more skilful will be the evaluation of the evidence from imperfect sensors. The ship must be guided through water that can never be touched.

The sections above outline the history of empiricism in psychology and show some of the varied evidence adduced in favour of this paradigm. But the various approaches that have been guided by empiricist or constructivist assumptions have been sketched only in bare outline (except of course Brunswik's, which, although written by someone calling himself a functionalist, clearly makes assumptions of a constructivist nature). It is time to give a more detailed account of one modern version of empiricism, a contemporary theory of perception that can be traced back to the work of earlier psychologists such as Bruner, and beyond them to Helmholtz himself. The theory of perception developed by the British psychologist, R. L. Gregory, will demonstrate where empiricism has arrived 100 years after Helmholtz introduced the idea into the psychology of perception.

Gregory's theory: perceptions as hypotheses

R. L. Gregory

Richard Gregory, now Emeritus Professor at the University of Bristol, England, is an experimenter of unusual originality. He has also invented a microscope, a telescope, a new type of hearing aid, and several other ingenious devices. In Gregory's experiments, observers have been hurtled down tunnels, swung on giant swings, and baffled by illusions. His lectures are distinguished by the use of novel demonstrations that are so compelling that one comes away convinced that what he says about perception must be right. He has inspired the building of a 'hands-on' science fair, The Bristol Exploratory, and is a frequent broadcaster. The quality of his writing and his demonstrations matches the standards set by the Gestalt psychologists and

Ames and his co-workers – which is to say, they are as good as any in the history of perception. Through his lectures and books, Gregory has enabled countless individuals to experience some of the delights that the study of perceptual phenomena affords. It is hardly surprising that his views on perception should be so well known.

Perceptions as hypotheses

As well as being well known as a highly original experimenter, Gregory is interested in the classical philosophical problems associated with perceiving. In an article published as part of a debate between psychologists and philosophers (Gregory, 1974), he describes some of the properties of perceiving that, he claims, force the conclusion that this is an activity resembling hypothesis formation and testing.

The essence of Gregory's hypothesis theory is this. Signals received by the sensory receptors trigger neural events. Appropriate knowledge interacts with these inputs to create psychological data. On the basis of such data, hypotheses are advanced to predict and make sense of events in the world. This chain of events is the process we call perceiving.

One of the merits of what Gregory calls the *hypothesis theory* is the clarity of its presentation. Another is the care with which Gregory presents the evidence in support of his views. Gregory's main arguments for this recent version of empiricism will now be summarized (see Gregory, 1980a, 1980b, for the full version of these arguments):

- Perception allows behaviour to be generally appropriate to non-sensed object characteristics. We respond to certain objects as though they were tables, having, that is, four legs and rectangular tops, even though all we can 'see' are three legs and the trapezoidal projection of the top. Are we not using more than just sensory inputs to achieve these percepts?
- Perception can, in familiar situations, mediate skills with zero time delay. In a typical tracking experiment, the observer is asked to keep a pointer aligned with a moving target. This would seem to be an essentially visual task, and visual processes are known to require a finite time. However, if the target position in a tracking task is made regular and predictable, then the observer will be able to track the target with zero time delay. How is this possible without a degree of anticipation entering into the perception of the target?
- Perceptions can be ambiguous. Look for a moment at Figure 5.5, the Necker cube. As one stares at this familiar figure, its orientation may suddenly change: it is unstable. If a single physical pattern can induce two different percepts, perception cannot be tied to the stimulation in a one-to-one manner. In a related manner, the simple drawing in Figure 5.6 can represent two different shapes: is it a duck or a rabbit?

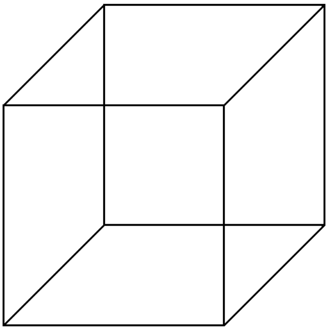


Figure 5.5 A reversible figure: the Necker cube. When fixated at the centre, the orientation of the cube may change quite abruptly.

- Perception can extract familiar objects from background clutter. Gregory uses as an example our ability to extract one person's voice from others in a crowded room. This is something that no machine has ever been able to do, although we find it relatively easy. Is this because there is a limited repertoire of acceptable speech sounds that we and the speaker share? If so, our knowledge of the language is reducing the informational demands of the task. The achievement that this form of perceiving represents is more obvious when we consider experiences in a foreign country: after a few repetitions, it slowly dawns on one that, for example, the sound, 'tootal urr', is the phrase, 'tout à l'heure'. Perceiving appears to be aided by knowledge.
- Highly unlikely objects tend to be mistaken for likely objects. One of Gregory's best-known demonstrations involves a hollow mask of a face. When viewed from the rear, the mask is generally seen as normal, with the nose pointed outwards towards the observer; in fact because the mask is being seen from behind, the nose is actually pointing away from the observer. Interestingly, even when one knows the true orientation of the mask (and even when one has constructed it) the illusion remains,

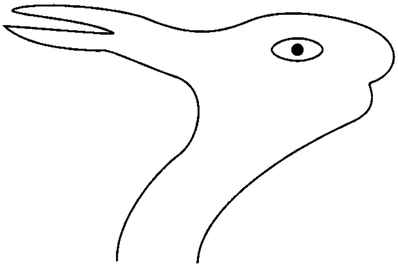


Figure 5.6 Duck or rabbit?

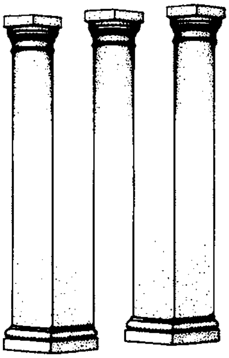


Figure 5.7 Shepard's intriguing version of an impossible figure entitled 'Doric Dilemma'. (From *Mind Sights* by Roger N. Shepard. Copyright © 1990 by Roger N. Shepard. Reprinted by permission of Henry Holt and Company, LLC.)

recalling Helmholtz's description of unconscious inferences as 'irresistible' and Brunswik's phrase, 'the stupidity of the senses'.

- Perception can be paradoxical. Figure 5.7 is Roger Shepard's version of a famous 'impossible figure'. Note how very difficult it is to arrive at an unambiguous perception of the figure, in particular to see how the central column is supported, given the changing status of the two central vertical lines as one fixates first the bottom and then the top of the figure. Shepard names his figure, 'Doric Dilemma'. The name seems very appropriate.
- Perception can be of one thing representing another. The perception of any picture is in a sense ambiguous: we see the lines and the surface and also the object depicted, even though the latter may in reality be many times larger than its depiction. It follows that there must be a large cognitive component in the perception of pictures.
- Perception is not essentially based on what is experienced. In many experimental situations observers may be influenced by stimulus characteristics of which they are unaware. There is a very simple and reliable way to demonstrate this. If one displays two photographs of a person, both printed from the same negative but in one of which the pupils have been enlarged, this will be seen as the more attractive version. Interestingly, observers are often unable to notice any physical difference between the two pictures: they have been influenced by the artificially dilated pupils, but have not noticed them.[4]
- People experience hallucinations. Quite simply, we can have vivid experiences of a perceptual nature in the absence of external stimulation.

---

[4] In medieval Italian states, women would put a plant extract containing atropine into their eyes to dilate the pupils. The name of the plant? *Atropo belladonna* ('beautiful lady').

These, then, are some of the key reasons that Gregory advances in support of the idea that perception is an indirect, constructive, hypothesis-like process. His case is clearly a formidable one. Notice also how well supporting instances are demonstrated: the statements above concerning empiricism and perception are some of the most explicit ever.

An application of hypothesis theory

Gregory has used hypothesis theory to develop an explanation of certain well-known illusions (Gregory, 1963). The Müller–Lyer illusion, which appeared in the present account of the Gestalt theory (see Figure 2.7), is one of the most famous of all the geometric illusions. For many years psychologists have attempted to explain the shortening and lengthening of the main parts of this figure, without, it must be said, much success. Gregory's suggestion is an interesting application of hypothesis theory to an old and hitherto intractable problem.

Suppose that the feather ends of the Müller–Lyer lines are acting as cues to depth, following the rule of linear perspective that parallel lines appear to converge with distance. By this rule, the shaft with the out-turned feathers triggers the hypothesis that there is an inside corner formed by the junction of two surfaces. Look now at the ceiling of the room in which you are reading this and note how the corner formed by the two walls and the ceiling resembles the out-turned feathers. Then inspect an object that has a corner or edge pointing towards you (say a book standing vertically on a horizontal surface), the outside corners of the book can be seen to resemble the in-turned arrows of the Müller–Lyer figure. In the real, three-dimensional world, such inward- and outward-facing corners reveal whether an edge juts towards or away from us, that is to say, they are cues to distance.

In real scenes, the shrinking of the retinal image of a receding object is opposed by the mechanism of size constancy, which, by enabling us to see things as the same size despite changes in distance, helps us to perceive a stable world. But if the Müller–Lyer arrows trigger this constancy mechanism, they are doing so in an inappropriate situation: the lines are actually equidistant from us on the page. So, instead of the equality of the shafts being preserved over different distances, the constancy scaling mechanism induces a perception of unequal size at a fixed distance, and this is the Müller–Lyer illusion: an adjustment of perceived size triggered when it is inappropriate. This original and ingenious explanation follows quite naturally from the assumption that perceptions are hypotheses, and is a good example of Gregory's ingenious deployment of his theory. It must be stated that the Müller–Lyer illusion must in fact be more complex than this, as it works when the inward and outward angles are replaced by semi-circles and straight brackets.

An evaluation of Gregory's theory

Gregory's theory that perceptions are hypotheses is the most explicit and fullest development of the empiricist paradigm. We shall now offer some criticisms of the theory, some general, others more specific.

The nature of perceptual hypotheses

One possible criticism of Gregory's theory is this: if perceptions are hypotheses, what sort of hypotheses are they? In one formal approach to the philosophy of science it is held necessary to abandon a hypothesis when a single contradictory fact appears (see e.g., Popper, 1960). However, it is commonly accepted that this is too arid a view and does not represent how scientists actually behave. Neither do perceivers: we do not mistrust our senses following exposure to a single illusion. Scientists modify and elaborate hypotheses according to their success, or lack of it. But how do perceivers modify their hypotheses? Is this done according to the frequency of positive and negative tests, or is the modification based upon the strikingness of confirmatory and invalidating experiences? For example, learning that a photographed face is that of a mass-murderer certainly seems to effect a permanent change in one's perception of the face. And the reader will have experienced rapid perceptual learning when discovering the face in Figure 5.2. On the other hand, learning to adjust to the effects of lenses which distort the world may take hours or even weeks of exposure. What is the difference between these forms of learning? Is it a difference between learning and experience – are the two different?

Hypotheses and language

A related general criticism concerns the relationship between hypotheses and language. What is the nature of this relationship? Hypotheses, in the normal use of the term, must be statable if they are to be tested against evidence. But we often have perceptual experiences that are hard to describe. It can be the case that only after considerable thought by the observer can he or she describe what was seen. But the seeing came first. It would appear that perceptual hypotheses might be closer to intuitions than to formal statements. We can say that we dislike someone without being able to say why. We can walk down a flight of stairs without looking at our feet and without even noticing what the stairs look like. The hypothesis that the stairs are regular does not appear in consciousness. Gregory is adamant that perceptual hypotheses are not represented in language – he uses the intriguing analogy of curve fitting:

> The curve fitting, though not necessarily touching experimental points on a graph, is a non-language of 'truth'. It interpolates, and may

> extrapolate beyond the evidence . . . so perceptual hypotheses are not unique in these respects.
>
> (Gregory, personal communication)

If hypotheses are not verbal or even conscious (remember Helmholtz and *unconscious* inference), then finding out about them is going to be difficult.

The inadequacy of sensory evidence

As has been stressed several times in this chapter, many theorists since Helmholtz have accepted his claim that sensory inputs alone are insufficient to specify the world. In support of this claim, it has been pointed out that, for example, because the retinal image of an object shrinks as the object recedes, the correct perception of the unchanging size of the object implies that the sensory evidence has been supplemented from other sources. This constancy example has been used several times in this book – quite deliberately, as it is a classic illustration of the empiricist argument. But are retinal images really so impoverished? In the world (in contrast to the laboratory) retinal images will only rarely contain projections of single isolated objects. They will be much richer than this, typically including projections of other objects, the background to the objects, even the distant horizon. They will be rich in detail.

Some modern research has shown that complex images of real scenes commonly contain information that can be used to tell whether a receding object has or has not changed its size. Basically, what seems to be important is that some part of an otherwise changing image remains *invariant*. Similarly, although the shape of an object may not be uniquely specified by any single view of it, multiple views may deliver an unambiguous and correct solution to the true shape. And movement is a vital part of perceiving, a truth which has often been overlooked in laboratory research. The search for invariants, the importance of multiple views, and the difference which movement makes to seeing will be discussed at some length in Chapter 6. For now it suffices to say that empiricists may have underestimated the richness of sensory evidence when perceivers operate in the real world.

Starting to perceive

This is a problem that seems at first to be rather trivial, but on reflection can cause complete bafflement. If perception is essentially constructive, how does it ever get started? How does the naive, newborn perceiver ever establish a grasp on reality? One wonders whether perception can be such an individual and chancy process. One's real-life experience suggests a great communality among the perceptions of different people. Where did this come from, if all have had to construct their own idiosyncratic worlds? We shall offer a partial answer to this question later, when discussing modularity.

Another partial answer is contained in the work of Brunswik (described in Chapter 3). Brunswik would stress that the selection of appropriate cues is vital to survival, and that organisms that get things wrong are unlikely to survive. That is to say, the world is common to all perceivers, and it may be this single fact that regularizes the perceptions of all creatures sharing a particular ecological niche. But hypothesis theory does not engage in such functionalist explanations.

The next general criticism of hypothesis theory may be somewhat unfair, in that a theorist is not obliged to consider all possible ramifications of a theory; what follows should be read as a statement of opinion rather than a formal criticism.

Human and non-human perceivers

There is a trap awaiting all whose work concentrates upon the human perceiver: it is to suppose that all perceivers are like us. But all species are the product of a long evolutionary history that shaped the structures *and* the functions of the senses. It is as well to remember that, on several criteria, humans are not the most successful creatures to date. They are certainly not the most numerous, neither does their history match the duration of other groups, such as dinosaurs and arthropods. All animals perceive, and must do this well enough to survive.

When we learn of the perceptual abilities of other creatures, we find much that is strange:

> The mayfly lives but a day as an adult. It may, for all I know, experience that day as we live a lifetime.
>
> (Gould, 1980)

Our (perhaps understandable) self-centredness should not blind us to the remarkably different lives of such organisms. Has the mayfly time to form and test hypotheses? Has it the neural equipment to do the necessary statistical assessments? Probably not. But mayflies can see their world, and have been doing so well enough to survive for millions of years.

There are certainly major qualitative differences between our perceptual systems and those of many other animals. The long period of postnatal helplessness in humans may be the price paid for perceptual flexibility, and in this they differ from many other species who can function well at birth but who are relatively inflexible in their subsequent behaviour. But at what point does perception cease to be reflexive and become constructive? In how many species can we apply the theory that perceptions are hypotheses? Might it be that our long evolutionary history enables us, too, to perceive the world more directly and automatically than hypothesis theory suggests? A partial answer may be contained in the proposal that perceptual input systems are *modular* – a point to which we shall return below. There will be more to say about this general problem in a later description of the theory of direct perception (Chapter 6).

Hypotheses and evidence

If we accept for the moment the idea that perceptions are hypotheses, a little thought reveals another serious problem: what is the *evidence* against which they are tested? Gregory is not clear on this point; it is, admittedly, a difficult one. One of his own examples can be used to show the problem. He says, rightly, that we frequently 'see' a table when its retinal image must be distorted and incomplete. We go beyond the partial evidence of our senses via the (reasonable) hypothesis that there is a rectangular table in view.

Such a table is in view as this is being written. The top is built from parallel rough planks (university salaries being what they are), and three legs are visible. It would be a shock to discover that this familiar, shaky object was not a table: the hypothesis seems to be a strong one. But what is the nature of the supporting *evidence*? Presumably it lies in the perception of the planks and the legs. But what guarantee is there that actual legs and planks are there? Is it not necessary first to have hypotheses to 'acquire' these components; and doesn't this lead to a regress of hypotheses concerning finer and finer details of the world? But if the Gestalt psychologists were right, the parts of a table do not simply add up to give the whole: they are seen in a manner partially determined by this whole. This and related criticisms of Gregory's theory are discussed more fully by the philosopher G. E. M. Anscombe (in Brown, 1974). Anscombe reminds us that a hypothesis is typically something that is answerable to data. What are the data to which Gregory's hypotheses are answerable? There may in fact be an answer to this problem. It has been provided in the work of Fodor (1983). In a now famous and highly influential book, *The Modularity of Mind*, Fodor seeks to revive and develop a very old idea, namely that there are mental 'faculties'.

Fodor argues for an important distinction between the mind and perceptual input systems. The more 'central' systems, the functioning of which give rise to mental phenomena (consciousness, awareness, thought and problem solving), are essentially 'horizontally' organized. They are unencapsulated and global in nature. An example will help flesh out this idea. A striking property of human thought is the ability to reason analogically. This appears to be particularly true in the development of new art forms and in the making of dramatic scientific discoveries. Fodor shows how things get likened to other, very different, things. The solar system is suddenly seen as a model of the atom; the benzene molecule as a snake-like ring. Insights play important roles in the history of discovery and invention.

In contrast, there are the perceptual input systems. The essence of these, Fodor argues, is that they are modular. By this is meant that they are self-contained, have limited tasks to perform, are reflexive in nature, and are cognitively impenetrable. For example, it is simply impossible to open one's eyes and *not* see a red surface as red. No amount of thought, no strongly held belief, no effort of will can alter such a basic visual response.

More controversially, Fodor includes language in his list of perceptual modules. In a similar argument to that above, he claims that it is impossible for a native speaker of English to hear the spoken language as merely sounds: the module will operate, reflexively, no matter what mental set the listener adopts.

It should be stressed that Fodor's reasoning is much more detailed and thorough than this short précis might suggest. For now, though, the thing to stress is the idea that perceptual input systems are by nature modular. However, this does not imply that they are merely automatic transducers of stimulation. They have a very complicated job to do, which is to represent information about the world to the mind (brain) in forms that it can use. If this is true, then we can begin to understand the nature of the evidence that Gregory's hypotheses are answerable to: the automatic outputs from perceptual modules. Of course, this conclusion is not the final answer to the question of how much of perceiving is constructive in nature, but it does suggest directions for future research. A possible beginning might be a taxonomic classification of all those aspects of perception that experience, mental set, and so on can (or cannot) influence.

This account of Gregory's theoretical work has concentrated upon the best known of the writings in which he defends his proposition that perceptions are hypotheses. In fairness, it should be mentioned that Gregory's views continue to develop and that he has refined his theory in his later work (Gregory, 1995). Further, in a personal communication, Gregory has informed the present author that he has read the previous edition of *Theories of Visual Perception* and that he has some points to raise. For the author of a textbook on theories, it is a rare privilege to receive comment from one of the theorists whose work has been evaluated. In reply to the comments and criticisms made above, Gregory's replies would be as set out below.

First, Gregory acknowledges the importance of 'bottom-up processes' – most evident in reflexive behaviour, such as blinking the eyes to a looming object. No amount of knowledge allows us to modify this response. Even experienced weapons instructors blink at the sound of a gun being discharged. And much of our behaviour towards objects goes beyond their simple optical properties: we can tell how and where to grasp them, for example. In order to be able to do this, we must be using internal representations in a 'top-down' manner, which in turn allow our intelligence to do its job.

Further, using conceptual knowledge in a top-down manner may take time. But perception is commonly very fast – there may be insufficient time for conceptual knowledge to play its part. We can witness this happening when, for example, we still experience illusions that we understand and are familiar with (remember Brunswik's 'stupidity of the senses'): the percept is formed before intelligence or knowledge can be used to get things right. Gregory therefore proposes that, in addition to 'bottom-up' and 'top-down' processes, there is another stage in visual perception. He calls this stage, 'sideways floppy disk operating rules'.

The importance of the sideways stage in Gregory's theory is this. There are many perceptual situations we need to deal with that are not handled adequately by reflexes (eye blinks and so on), neither can they be handled by our acquired knowledge of the world. For example, during a person's first-ever exposure to a pair of stereograms in a stereoscope, it is well nigh impossible for them to know what they 'should' see. But most people will eventually experience a sensation of depth. If this is due neither to a pure visual reflex nor to an intellectual solution to a problem, what is mediating this unfamiliar perceptual experience?

The answer to this question might be that, in normal binocular viewing, the visual system must in some way compare the two retinal inputs in order to use differences between them as cues to the relative depths of objects (a topic that will be dealt with at greater length in Chapter 7). To do this requires specific procedures, of which we are of course unaware – they are not part of our conceptual knowledge. If we understand Gregory's position on this issue, the stereoscope situation would be one where the specific procedures are inserted sideways, as it were: the floppy disk operating rules. Thus, something is being added to raw sensory inputs – which is the essence of the constructivist position.

There will be times when the sideways disks will not be available, or when an inappropriate one has been selected. For example, it might be because we insert a disk containing procedures for dealing with objects and not pictures that we fall prey to perspective tricks induced by the latter.

Gregory raises many other interesting points concerning visual perception, particularly those associated with what he considers to be the ill-judged analogies drawn between perceivers and machines. Some of these will be included in later parts of this book.

Concerning hypotheses and induction, Gregory raises the important point that although Popper rejects induction as the basis of science, in fact one needs induction in order to be able to trust that evidence one might use when abandoning a hypothesis. This is a very interesting paradox.

Gregory admits that we really do not know how perceptual hypotheses are modified.

With regard to the stubborn persistence of illusions despite familiarity and knowledge, Gregory emphasizes the fact that he has always drawn a distinction between perceptual and conceptual hypotheses. Perceptual and conceptual learning are not the same – there is only a slim connection between them. Perceptions must work very quickly if they are to help the perceiver survive. In contrast, the search for conceptual solutions to problems can take months or years.

Gregory concedes that perception 'gets off the ground' in children via the action of certain 'innately significant shapes' such as faces, and certain key features indicating form, including the perception of corners and so on. Significantly, such key features appear to play important roles in eliciting distortion illusions, such as the Müller–Lyer pattern discussed earlier.

Finally in this personal correspondence, Gregory offers a striking speculation, namely that there may be a deep connection between perceptual hypotheses and language (not that perceptual hypotheses involve language in Gregory's view). The speed with which language developed in human history suggests that it derived from existing perceptual structures, namely very ancient perceptual classifications of objects and possibilities of action.

Computational theory and top-down processing

In Chapter 7 an account will be given of Marr's computational theory of vision. Much of this modern research, as we shall see, concentrated upon the ways in which information is extracted from the visual image on the retina. But Marr was mindful of the fact that analysis must be followed by synthesis, and he acknowledged the role that knowledge can play in contributing to such a synthesis, although he may have underestimated the importance of this role. However, it is clear that empiricism is alive and well.

Final remarks on empiricism

The theoretical writings that have emerged within the empiricist paradigm have tended to use psychological rather than physiological concepts. Gregory's theory, for example, is closer to Brunswik's than to the Gestalt theory. Earlier in this book (Chapter 4), we have outlined an argument that believes this to be quite proper: problems in the psychology of perception demand explanation at the appropriate level. Pain is something we feel. Although it is undoubtedly caused by neural impulses, these are not part of our awareness. Pains may be sharp or dull; neural impulses, as such, are neither.

The tradition (or paradigm) that has been outlined in this chapter has been a vigorous one. A mass of results has been obtained from highly ingenious experiments. Very little of the literature in the area can be dismissed as trivial or dull – quite the reverse, as anyone may confirm by reading, for example, Gregory's publications. That words may affect us below the threshold of awareness is a strange fact. The effects of set and attention are fascinating, as is the fact that we can be so completely fooled by an oddly shaped room or a hollow face. It is a rewarding experience to introduce people to such phenomena, as any teacher of perception will confirm. And, as has just been stated, the empiricist or constructivist approach is still to be seen in perceptual theorizing.

Early in this chapter an account was given of some of the discoveries that inspired empiricism. Later, some criticisms were put forward concerning Gregory's hypothesis theory. At this point it seems reasonable to offer a general opinion about empiricism as a paradigm for perception.

First, it must be said that nobody is yet in a position to make a final judgement between constructivist and rival approaches to perception. The deep mysteries of perception remain and it requires an act of faith to believe that they will ever be solved. What follows is a speculation of the kind that must have occurred to many who have tried to evaluate empiricism.

The evidence adduced in favour of constructivist accounts of perception, such as hypothesis theory, comes in the main from one of two types of experimental situation: the stimuli employed are either meaningful, in an abstract sense, or they are products of the built or cultural environment. Consider: the perception of patterns under conditions of brief exposure; drawings that could represent the corners of buildings; oddly shaped rooms; hollow masks; twin-track tape recordings; glowing objects in darkened corridors. These are the sorts of situation faced by observers in many of the classic experiments that have sustained empiricism. But none of these existed in the African grasslands, where human perceptual systems reached their present state of evolutionary development. The evolution of the modern human being obviously antedated human civilization. Has this research been appropriate?

This general point may be reinforced by another example. Consider the problem of flying. Most people could be taught within minutes to keep a light aircraft straight and level under conditions of good visibility. In fact, people have stolen aircraft and taken off successfully without ever having handled the controls before (most of them died when attempting to land, however). But nobody can fly for long in cloud without special instrument training – a claim that is borne out by a long list of fatalities. Why is this? The reason is that when an aircraft starts to deviate from its heading, for whatever reason, detectors in the inner ear correctly signal the initiation of the turn. But as the turn continues, the lack of change of radial acceleration causes these same detectors to signal that the body is now travelling straight ahead. At this point any attempt to straighten the aircraft will feel like a turn in the *opposite* direction. The situation is now out of control and can only worsen. The pilot, who cannot get back into step with the manoeuvres of the aircraft, is about to become an accident statistic.

For a trained pilot, the situation is completely different, and quite safe. He or she has practised ignoring sensations from the inner ear in order to concentrate upon the readings from the flight instruments. In time, these seem to become the 'natural' source of information about the behaviour of the aircraft. But this takes much learning. And it is of course highly artificial and almost completely cognitive, at least in the initial stages of training.

It is hard to resist the conclusion that perception under blind flying conditions is learned, interpretative, and constructive. And this may be true of perception whenever the situation is in any way artificial or unnatural. This must happen whenever meaning must be extracted from a symbolic display. Meaning clearly implies knowledge, but a word does not signal directly what it stands for: ‘fin’ is part of a fish in English but means ‘end’ in French; ‘Adler’ is a gentle word in English but means ‘eagle’ in German.

Hence it is possible that we can perceive constructively only at certain times and in certain situations. Whenever we move under our own power on the surface of the natural world, and in good light, the necessary perceptions of size, texture, distance, continuity, motion, and so on may all occur directly and reflexively. The claim that this is in fact the case is a tenet of the theory to be discussed in the next chapter, which will end with an attempt to form a compromise between the empiricist/constructivist approach and one of its main rivals: the theory of direct perception.

The relative brevity of this chapter should not be taken as an indirect evaluation of the importance of constructivist explanations of visual perception. On the contrary, it would be quite possible to write an entire volume on this topic alone. This account has been kept as short as possible for two reasons. First, the empiricist approach dominates much modern thinking. The reader has only to consult any standard general perception text to find all the main empiricist demonstrations and interpretations stated with clarity and conviction. It can be claimed that this paradigm has been so dominant during the past half century that, until recently, it *was* the general theory of perception. Second, when describing a radically different approach to perception in the next chapter, it will be necessary to describe again many of the claims made by modern empiricists. So we have not yet finished with the general topic of empiricism or the specific version of it represented by Gregory’s hypothesis theory.

Our final conclusion regarding constructivist theories of perception is as follows. If perception is neither a set of capacities fully determined at birth, nor completely learned during life, then in terms of theory there must be something analogous to a pendulum that can swing between these two extremes. In our view, this pendulum is currently swinging away from the empiricist position. The evidence showing a remarkable degree of perceptual competence shortly after birth (described in Chapter 2) is very striking and must surely force empiricists to retreat somewhat.

Does this mean that a theorist working from an empiricist (or constructionist) point of view will shortly be out of a job? We think not. Remember the tragic case of S.B. and how he was unable to recognize a lathe until he touched it – then the visual image of the lathe suddenly acquired *meaning*. Something had changed the perceptual input and rendered it comprehensible. Philosophers have written about the difference between seeing a star and seeing it *as* the Pole star, and this describes the phenomenon succinctly.

The programme of learning about how we perceive the meanings of the world is already under way. Those whose work has been described in this chapter have commonly focused their attention on problems such as how meanings can distort geometric figures, how pictures and symbols come to be interpreted as representative of other things, how we learn to recognize others' faces, and how we interpret ambiguous displays. Given the size and complexity of the social/cultural environment in which we live, psychologists will be kept busy for many years yet.

Endnotes

- Readers wishing to learn more about the history of the nativist/empiricist debate referred to in this chapter, and in Chapter 2, should consult Gordon and Slater (1998).

- Helmholtz's position relative to the 'inference revolution' is clearly described in Gigerenzer and Murray (1987).

- For a philosopher's reaction to the claim that perceptions are hypotheses, see the comment by G. E. Anscombe published in Brown (1974).

- Constraints of space precluded any discussion of the work of Rock (1995), another modern worker to have included aspects of empiricism into his theory.

- We suggested to readers that the wire from a champagne cork could be used to demonstrate reversed depth and movement. Another fascinating effect can be induced by illuminating a hollow face mask obliquely at the rear. The nose will now protrude; when the mask is rotated slightly, the movement of the nose will be in the opposite direction to the rotation of the mask. The present writer first experienced this intriguing illusion during a lecture given by Richard Gregory.

- In a personal communication, Professor Gregory has informed the present author that in fact Helmholtz demonstrated the Ames room 90 years before Ames in a popular lecture. Perhaps the demonstration should be named the Helmholtz–Ames room?

6 Direct perception and ecological optics: the work of J. J. Gibson

> . . . perceiving is an act, not a response, an act of attention, not a triggered impression, an achievement, not a reflex.
>
> (Gibson, 1979)

The theoretical position to be described in this chapter owes a great deal to the work of one man, the American psychologist, J. J. Gibson. His claim that perception is in an important sense direct, and his development of what has been called ‘ecological optics’, are among the most interesting theoretical developments in modern perceptual research. Since his death, Gibson’s ideas have been refined and developed and he himself changed his views during the course of his career. In what follows we shall give a general account of what seem to be the most important aspects of this approach to perception; for the sake of clarity and economy, we shall not always indicate whether a particular idea or argument belongs to Gibson or to a follower of his, although major theoretical differences will be pointed out. The general term, ‘direct perception’, will be adopted. This has been given to the body of theory developed by Gibson and his followers that, it has been claimed, represents a new paradigm. The reader will note that, once again, visual examples dominate the account of a theory.

The remainder of this chapter will cover the following topics:

- J. J. Gibson.
- An outline of the theory of direct perception.
- An evaluation of the theory of direct perception.
- More recent research.

J. J. Gibson

Gibson was born in 1904 and died in 1979. He was educated at Princeton and later took a teaching post at Smith College. He became known for his experiments and his theoretical writings after moving to Cornell University, where he stayed for the rest of his career.

Gibson’s education gave him, initially, a behaviourist approach to his subject, although by the 1960s Gibson had come to disagree fundamentally with the assumptions of behaviourism. In fact, as his friend and colleague R. B. MacLeod has pointed out (MacLeod & Pick, 1974), in one sense Gibson was a functionalist of the old pre-behaviourist school. It must also be pointed out that Gibson came into contact with the distinguished Gestalt psychologist, Kurt Koffka, towards the end of the latter's career, and came to hold his work in high esteem.

As a young experimental psychologist, Gibson worked on a variety of problems. He was interested in the effects of mental set on performance, he studied human conditioning, and he did orthodox psychophysics. He was then an empiricist – a theoretical position that he gradually abandoned after studying adaptation effects in perception. It was known that if an observer wears spectacles that distort the visual world, prolonged exposure to the distortion leads to a degree of recovery. For example, if the spectacles cause vertical lines to appear curved, the lines seem to straighten after a period of practice. Removal of the spectacles then causes the world to bend in the opposite direction for a time. The usual explanation of this ability to adapt to distortion was that the brain gradually comes to reduce the discrepancy between the distorted visual input and normal tactile inputs: in Berkeley's original sense, touch teaches vision.

However, Gibson found (to his surprise) that adaptation to the spectacles occurred if the observer simply sat and stared at vertical lines. Further, simply staring at curved lines, without using spectacles, caused their curvature gradually to lessen. Such effects convinced Gibson that perception could not be merely a compound of simple sensations and that empiricist interpretations must be flawed. Much of the remainder of Gibson's career was devoted to attacking what he considered to be the misleading and harmful notion of sensation.

During the Second World War Gibson worked on the applied problems of pilot selection and testing. Flying clearly demanded accurate perception of space, but:

> . . . as I came to realise, nothing of any practical value was known by psychologists about the perception of motion, or of locomotion in space, or of space itself. The classical cues for depth referred to paintings or parlour stereoscopes, whereas the practical problems of military aviation had to do with takeoff and landing . . .
>
> (Gibson, 1967b)

Gibson became convinced that perception by pilots of aircraft made important use of information from the ground and the sky (particularly when there was a covering of high cloud), and that this information was in the forms of patterns of movement, the flowing textures that arise as a result of motion relative to the ground (and sky – when conditions permit). Gibson's preliminary analysis of this situation is shown in Figure 6.1.

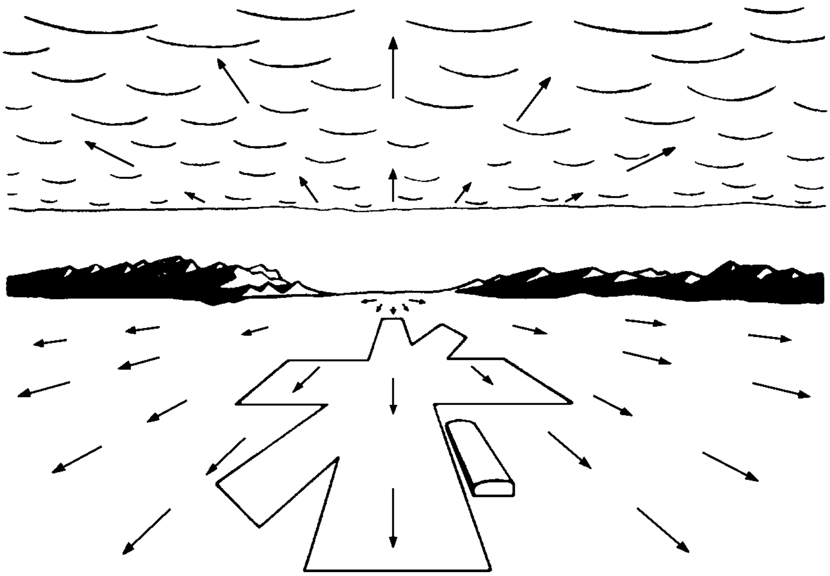


Figure 6.1 Optic flow during a landing approach. This is the sort of visual phenomenon that was brought to Gibson's attention during his involvement in flying training. (From Gibson, James J., *The Perception of the Visual World*. Copyright © 1950 by Houghton Mifflin Company. Reprinted with permission.)

In 1950 Gibson expressed these views in detail in his book, *The Perception of the Visual World*. This became a classic and its main findings are now included as standard in most textbooks on the perception of space. The importance of movement in perception and the usefulness of considering perception under real-life conditions, as opposed to simple laboratory experiments, were beliefs that remained with Gibson for the rest of his professional career. Gibson's theoretical position evolved over the years. The following account will tend to emphasize his later ideas.

With the possible exception of his work on perception and art, all Gibson's writings are original and interesting. He wrote superbly and is still well worth reading.

An outline of the theory of direct perception

Objections to empiricism

A good way to appreciate the arguments for direct perception is to understand what Gibson and his followers objected to in the most popular contemporary paradigm for perception, namely empiricism. We have outlined this approach in Chapter 5. As a reminder, here is a summary (some would say a parody) of this position.

In any momentary visual fixation of the world, the relationship between distal and proximal stimuli is likely to be imperfect: retinal images shrink as objects of fixed size move away from us; a table shows only three legs; tilted rectangles yield trapezoidal retinal images. We see things that are not physically present when we complete gaps in patterns or see illusory contours. Colour resides not in objects but in our heads. The sensation of tickle does not resemble the objects that induce it. In other words, sensory inputs are commonly too impoverished or too degraded to specify aspects of the world.

Because sensory inputs (or the resulting sensations) are not rich enough to mediate perception, the perceiver must add to them. The elaboration of sensory data involves inferential processes utilizing memory, habit, set, and so on. Survival pressures require that inferential processes deliver 'correct' solutions most of the time – we successfully go beyond the sensory evidence – but sometimes inferences fail and we experience illusions or other 'errors' of perception.

The essence of this constructivist paradigm, therefore, is that perception of the world is essentially indirect: something must be added to the incoming stimulus information before the final perceptual response is attained; sensory inputs must be represented as images, schemata, or models.

Gibson and his followers argued that this assumption leads inevitably to a particular research strategy: if the visual image is the starting point for elaboration, study visual images. If successive samples of the world are important, present such samples under controlled conditions using brief exposures. To present brief exposures in a controlled manner, keep the viewer's head still. The use of brief exposures will eliminate errors due to eye movements. And so on.

Data from such studies must then be fitted into some sort of model. As events take place 'in' time (the time of Newtonian physics – even, unbroken, unidirectional), the perception of events includes the perception of their sequence and of time itself. Thus, the model chosen for perception will inevitably involve stages: successive samples must be stored before being elaborated. This in turn requires the involvement of different types of memory: iconic, short-term, long-term, and so on. And as the model now includes stages, it is natural to think in terms of information flowing between them. Inevitably there will be the conceptual leap into believing that perceptual processes resemble the workings of the *digital computer*.[1]

As we attempted to show in the previous chapter, the constructivist or empiricist approach has been a fruitful way of thinking about perception. It has generated numerous ingenious experiments, yielding important data. But the question remains: is awareness only indirect? Is our perceiving really mediated by internal representations? Direct perception theorists think not.

---

[1] This was true when Gibson was first developing his theory. Gibson died before connectionist networks began to make an impact in psychology (see Chapter 4).

Gibson and his followers (for valuable discussions of what follows see e.g., Costall, 1981; Reed, 1987) argue that the constructivist, indirect paradigm has a very long history (which explains in part why it is so pervasive). The Galilean doctrine that nature is composed of matter residing in physical space and time led to the Cartesian doctrine of the essential separation between the mental and the physical. This raised, inevitably, the major philosophical and psychological question of how the realms of the physical and the mental meet: if our minds are essentially different from the world, then we cannot know it directly; all we can know are images of the world – sensations arising from it that are used to represent it.

Reed (1987) points out that, to this day, psychologists tend to view space and time as the 'receptacles' of objects. There is, thus, an automatic tendency to separate psychological activity from the biological and physical aspects of the perceiver. Seen from this perspective, the physical world is meaningless and neutral. Gibson's aim was to find out how organisms become aware of this world, how they come to behave as though the world is sensible and meaningful. Hence, although he was against cognitivism (the postulation of mental representations formed from sensations, and so on), what Gibson attempted was a cognitive theory: he wanted to explain how organisms come to know the world. But in seeking this explanation, Gibson was determined to avoid the dualism inherent in the traditional view of the perceiver, which separates perceptual experience from the objective world. His work thus represents a radical challenge to the existing philosophical framework within which most theories of perception have arisen.

To achieve his aim Gibson was led to reconsider the nature of stimuli and the ecologies, and their relationships with perceiving organisms. We shall see that, by the end of his career, Gibson had arrived at a new way of describing stimulation; he had rejected sensations as useful explanatory concepts, he had abandoned the distinction between sensory and motor aspects of behaviour, and he had given a new impetus to the study of the environment and its inhabitants. More fundamentally, Gibson was able to claim that when the appropriate ways of describing perception had been found, many of the problems that had engaged earlier theorists evaporated.

As a starting point for an attack on the idea of indirect perception, Gibson and his followers would begin with a discussion of the nature of light.

Light and the environment

In any textbook discussion of, say, the problem of size constancy (to use a familiar example), the starting point is usually a simple optical diagram. Single lines, representing rays of light, are drawn from an object to the eye (see Figure 6.2). When the object is drawn as further from the eye, the ray diagram shows how the visual angle at the eye is diminished, as is the (inverted) visual image on the retina. Why then doesn't the object appear smaller to the viewer?

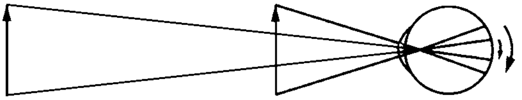


Figure 6.2 The reduction of image size with distance.

The problem changes, however, when we consider a real scene. The viewer is in, say, the centre of a room, with a light source overhead. The light source (a window, a lamp) is emitting light in many different directions. This emission may comprise several million rays, in contrast with the one or two in a classical ray diagram. Further, not all the light from the source comes directly to the eye: some rays (a few million) may reflect from a wall to the object and then to the eye; others may strike the floor and then the object before entering the eye. Other rays may come to the eye not from the object but from the surface on which the object is standing. The eye is bathed in a sea of radiant energy, of complex interactions between light rays moving in different directions, many of which have been reflected by surfaces. The visual world comprises surfaces under illumination.

The next point is so obvious that it is in danger of being overlooked: it is *because* light travels in straight lines that it can carry information about the environment through which it has travelled and from which it has been reflected. In a mad universe in which light rays swerved erratically, light could not be informative.

It is a happy chance that since Gibson started writing about the richness of light in this way, the development of the laser hologram has provided a powerful confirmation of his claim. When laser (that is, very coherent) light is shone onto a real scene, reflected rays can be captured on a photographic plate. If a reference beam is now shone directly onto the plate, the two beams form a complex interference pattern. When the plate is developed photographically and illuminated with laser light, the original scene is recreated in an extraordinary manner: it is in the form of a three-dimensional image that can be studied as though it were the original (see Figure 6.3). If one object is in front of another, one can move one's head and look behind it; one can focus a camera on different objects at different distances in the three-dimensional space. But when the photographic plate used to generate the hologram is scrutinized in order to find an image or picture of the original scene, none is to be found. What is there on the plate is simply the interference pattern. Moreover, the information necessary to create the hologram is contained all over the plate: one can break the plate into small pieces and each piece can then be used to generate a three-dimensional image of the original scene. Light (and interactions between light rays) can be a rich source of information.

To summarize: if we examine light arriving at the eye in real situations, we find that it is structured. It is highly complex and potentially rich in

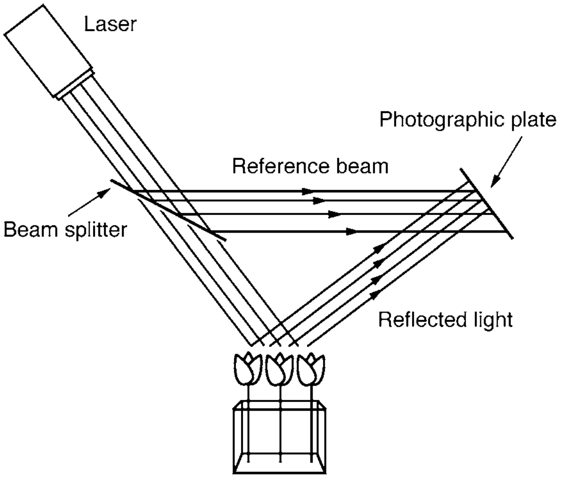


Figure 6.3 The construction of a hologram. The reference and reflected beams from the laser form an interference pattern on the photographic plate. When this plate is developed and illuminated with coherent light from a laser, a three-dimensional image of the original scene is generated.

information. A single momentary retinal image may indeed be impoverished, but this is not true of the arrays of nested solid visual angles through which the head and the eyes sweep during normal perceiving. As we come to understand more and more about these arrays and the potential information contained in their structure, the less frequently will we need to invoke supplementary, indirect processes in explanations of seeing.

Perception and evolution

Gibson was undoubtedly influenced in his thinking by an important book by Walls (1942). In *The vertebrate eye and its adaptive radiation*, Walls presents a mass of evidence to show how the astonishing variety of vertebrate eyes can be explained by considering the range of habitats or ecological niches that their owners occupy. Any extant animal is by definition successful, embodying as it does the result of millions of years of evolution. To understand what an animal's perceptual systems can do, we must consider the environment in which they evolved, for it is this environment that shaped the systems. We should consider the animal and its environment as two interacting systems:

> The words 'animal' and 'environment' make an inseparable pair. Each term implies the other. No animal could exist without an environment surrounding it. Equally, though not so obvious, an environment implies an animal (or at least an organism) to be surrounded.
>
> (Gibson, 1979)

> An animal is what it is given that its niche is what it is; an animal's wings, gills, snout, or hands describe that animal's environment. Likewise, a complete description of a niche describes the animal that occupies it.
>
> (Michaels & Carello, 1981)

These statements are worth elaborating (as Gibson, and Michaels & Carello, do in their books), for they describe an idea that can be a powerful stimulus to the imagination. Consider this very unusual environment: boiling sulphurous mud. This environment can be analysed in detail – its lack of oxygen, its acidity, the ferocious temperature. When these factors are combined, we have effectively defined the only creature that could inhabit such a strange niche: it is in fact a rare bacterium. When, on the other hand, the structure of a relatively 'simple' multicellular organism, *Taenia saginata* (a tapeworm) is examined, it can be seen to lack musculature, thus it cannot move far in its daily existence. It possesses none of the usual major sense organs. *Taenia*'s body comprises a long chain of segments, each quite flat in form and sheathed in a membrane that is not destroyed by weak acids but which permits absorption by osmosis. We are close to defining the only environment that could support such a creature: the large intestine of a warm-blooded vertebrate.

The case for ecological optics

The environmental niche determines the structure of an animal and its senses. To understand the animal's perceptual systems, it is necessary to consider the environment in which these systems evolved. But what is it exactly that we need to know? Gibson's advice (in the case of vision) would be unhesitating: find out about the patterns of light that arrive at the eye from the environment and ask what potential information about the environment is contained in these patterns. This is a first step. Later we can discover whether particular aspects of this information are or are not utilized in perception. But we must begin by examining the lit environment, and to do this we need a new science: ecological optics.

When we draw simple ray diagrams (e.g., Figure 6.2), we are using classical optics. This is a science that is neutral with respect to the viewer, and extraction of principal rays is a legitimate simplifying exercise. But, as we have shown, pondering over simple ray diagrams makes the problem of size constancy seem formidable. Similarly, simple physical measurement leads to puzzlement over the phenomenon of brightness constancy: why, for example, does coal look black on a summer's day when it can be shown that the light that it reflects is many times more intense than that from snow on a winter's afternoon? How do we continue to perceive the 'true' properties of these stimuli when simple measurement suggests that this should be impossible?

One answer, using an ecological approach, is that classical optical science ignores the complexity of the real environment. When an object moves further away from the eye, its visual image does indeed get smaller. But this is not the only change occurring in the complex pattern of light arriving at the eye. Most objects have textured surfaces and the grain of this texture gets finer as the objects recede. Objects obscure a portion of the textured ground against which they are seen. The further away an object is, the closer it will be to the horizon, and so on. And although the light reflected from a dark object under strong illumination may be quite intense, it will be less intense than light from more reflective objects present in the same scene. The important point is that objects are not usually seen in complete isolation. The optical array commonly contains far more information than that associated with a single stimulus object. The use of classical optics and an over-concentration upon laboratory experiments may cause us to overlook this important truth. (See Gibson, 1961, for a much fuller discussion of this point.)

The role of invariants in perception

One of the most important concepts in direct perception theory is the invariant. The emphasis on invariants may turn out to be Gibson's single most important contribution to psychology, and understanding invariants is the key to understanding the theory of direct perception.

Gibson frequently stressed the importance of movement in perceiving (see e.g., Gibson, 1966). Indeed, he insisted that the distinction commonly drawn between sensory and motor aspects of behaviour is an artificial one and leads to false problems, such as the question, 'Why doesn't the world move when we move our eyes?' For Gibson, the changes brought about as a result of our motor behaviour should be thought of as an integral part of the process of perceiving. We rarely receive a static, unvarying view of any object or scene. We move our head and eyes, we walk around the environment, things come into and out of view: perception is an active process.

Imagine that one was reduced in size to the point that one could get inside a vertebrate eye. What would it be like, down among the rods and cones of the retina? What one would *not* see would be part of an image, the edge of a static picture. Instead, one would see shimmering patches of light flickering across the retinal cells. At any moment the textures in the environment would project countless points of light into the eye. And as the eye moved, fresh patterns would sweep across one's position. The scene would appear kaleidoscopically complex, even chaotic. But this jumble of coloured patches of light is not in fact random. Among the patterns of change are lawful regularities – the movements of adjacent parts are correlated. An example will serve to illustrate this claim. One is approaching a textured surface. At each moment the patterns of stimulation from the environment are changing. But this change, although complex, is non-random. Photographic analysis of the scene (and even informal introspection) reveals that the changes in the textures seen (and thus the changes at the retina) follow patterns of *flow*. That part of the vertical surface with which one will eventually make contact remains stationary, although growing in apparent size. All around that point, one can notice a radial expansion of textures flowing around one's head (see Figure 6.4). The textures expand as one approaches them and contract as they pass beyond the head. And the situation we have described will be the case whenever we move toward something. In other words, over and above the behaviour of each texture element is a higher-order pattern or structure, and this is available as a source of information about the environment. In this case the flow of the texture is described as *invariant*.

Here is another example of the lawfulness that can be exposed when familiar situations are examined thoughtfully. How do we know that an object that has gone out of sight has not gone out of existence?

> Any movement of a point of observation that hides previously unhidden surfaces has an opposite movement that reveals them. This is the law of

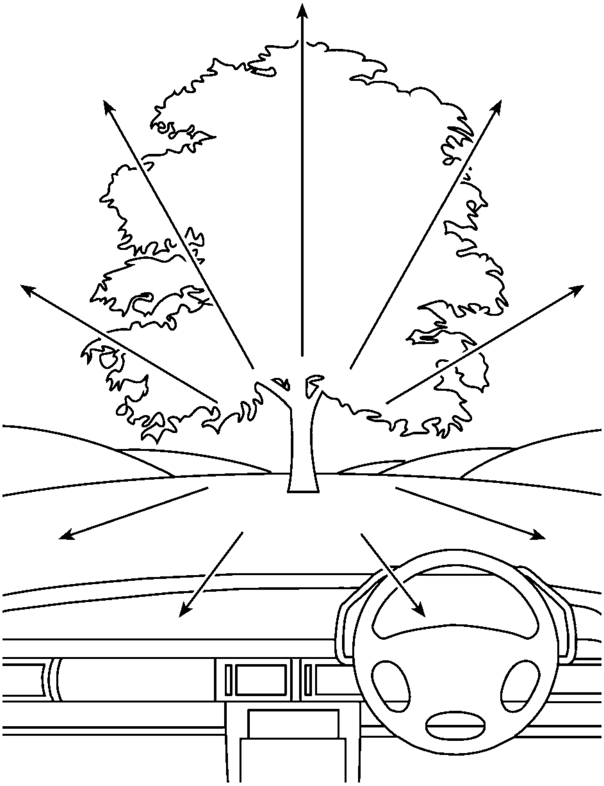


Figure 6.4 Optic flow surrounding the point of eventual collision.

> reversible occlusion, which states that the hidden and unhidden real things in a locale can be interchanged by moving around. Going out of sight is not the same as going out of existence. The perception of persistence does not rely on the persistence of perception, but on tests using reversible occlusion.
>
> (Reed, 1987)

The essence of invariants is that they are associated with change. They can be thought of as higher-order properties of patterns of stimulation that remain constant during changes associated with the observer, the environment, or both. Modern theorists distinguish between two types or styles of invariant, transformational and structural.

Transformational invariants

These are patterns of change that can reveal what is happening to an object or objects. For example, when an object moves away from us at constant speed, its apparent area (the size of the angle subtended at the eye) diminishes lawfully. In fact, the decrease in area is proportional to the square of the distance. Whenever this relationship is present, it must mean that the distance between us and the object is changing in a regular manner. Departures from this invariant rule can mean only that (1) the rate of movement has slowed or accelerated, or (2) the object is actually changing size. Here it is the *style* of change that is a source of information.

Structural invariants

These are higher-order patterns or relationships that remain constant despite changes in stimulation. A tune is a structural invariant: when it is played in a different key or by another instrument, the essence of the tune is preserved. As an example of a structural invariant in visual perception, consider a situation in which two objects having the same physical size are at different distances from an observer. Clearly, the visual angles subtended by the objects (and hence the sizes of their retinal images) will be different. How can we know that the objects are in fact the same size? This is of course yet another way of introducing the problem of size constancy.

Analysis of the situation described above reveals that there is indeed an invariant property of the stimulus array that could serve as information specifying that the objects are the same size. The invariant is a subtle one: if the objects are in a natural environment, then they will usually be viewed in a scene containing a visible horizon; it can be shown that the *ratio of an object's height to the distance between its base and the horizon is invariant across all distances from the viewer* (see Figure 6.5). Analysis of light with reference to the environment has yielded a possible solution to the problem of size constancy.

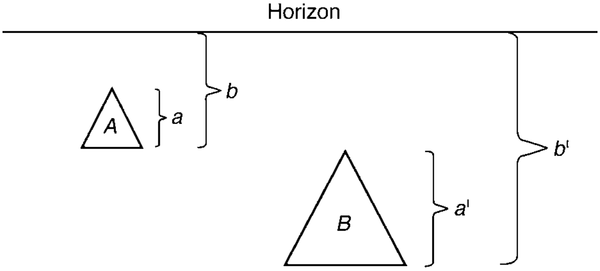


Figure 6.5 A possible invariant underlying size constancy. If $a:b = a':b'$ then *A* and *B* are the same size.

There is a similar regularity in the relationship between size, distance, and an observer's eye-height, which interested readers can check for themselves. Suppose your eye-height when standing is 150 cm. You are looking across a level view in which there are two trees, one near, one far. How can you tell whether they are the same height? The answer is that the horizon (because it is at optical infinity) will intersect each tree at a point 150 cm above the ground: your eye-height. If this is, say, a quarter of the way up each tree, *then the tree heights are identical*: 150 × 4 = 600 cm (6 m). What an encouraging demonstration of the richness of information that light can convey under natural conditions![2]

Affordances, invariants and meanings

In Gibson's later writings (e.g., Gibson, 1971a, 1977, 1979), increased emphasis is placed on the '*affordance*', a concept that has been refined by several of his followers. It is at this point in the theory that the relationship between perceiver and environment assumes great importance. The environment contains invariant information, the detection of which has survival value for a perceiver.

> Roughly, the affordances of things are what they furnish, for good or ill, that is, what they *afford* the observer.
>
> (Gibson, 1971a)

Gibson goes on to list a series of possible affordances. These include, for humans, surfaces that are stand-on-able or sit-on-able, objects that are graspable or throwable, objects that afford hitting, surfaces that afford supporting, substances that afford pouring. A single object may give rise to more than one affordance: an apple, for example, affords eating and grasping and throwing.

---

[2] Interestingly, it looks as though the horizon plays a similar role in the perception of distance, size and depth in pictures (see Sedgwick, 1980; Rogers & Costall, 1983)

It is clear that affordances are the *meanings* that an environment has for an animal. As meanings, the affordances guide behaviour: they tell the observer what is or is not possible. The range of possible behaviours in response to affordances has been described as the set of *effectivities* available to the organism, although some theorists believe that the term 'actions' is all that is required. It is clear that this part of his theory reveals the influence of the functionalist tradition on Gibson's thinking. The originality of Gibson's approach to invariants and affordances, these seemingly abstract properties of things and events, lies in his remarkable assertion that they can be perceived *directly*, without prior synthesis or analysis. Thus, the properties of an object that reveal that it is graspable (just consider the vast array of different objects to which this description could be applied) are there to be perceived directly from the pattern of stimulation arising from the object. This is a very bold idea.

Two further aspects of the theory of invariants and affordances should be stressed. First, we must remember that understanding perception requires the joint study of an organism and its environment. This essential relationship must always be borne in mind. As was stated above, when a piece of music is transposed to a new key, the new set of notes may be completely different. But something is preserved in this change, namely certain important relationships between successive notes. This identity (or near identity, for the situation is a little more complicated than this) of musical intervals provides a basis for the equivalence of tunes across keys. But this equivalence will be experienced only by a perceiver sensitive to interval information: it will be of little use to the tone-deaf. Similarly, when a rectangle is tilted away from us, its projected shape becomes trapezoidal (see Figure 6.6), but it continues to look rectangular. How?

There is a geometry that can treat a trapezoid as a *transformation* of a rectangle. That is to say, the rectangular shape can be 'recovered' from a trapezoidal projection. It follows that there is something invariant in the property of the shape that might allow an observer to decide that he or she is seeing a tilted rectangle. But whether or not sensitivity to this transformational invariant is the basis of shape constancy is an empirical question that is not solved simply by isolating the potential invariant: we need to know whether a particular perceiver can use such information. Gibson's important term here is '*attunement*': organisms need to be attuned to affordances before they can exert their power to shape actions.

Second, it is important to remember that an invariant or affordance for one species may not be an invariant or affordance for another. Sensitivity to certain odours, to ultra-violet light, or to the earth's magnetic field, provides some species with information that is quite outside our own direct experience, and which we could never use. Failure to recognize this important fact may be the reason for many misunderstandings of the apparent oddities of animal behaviour; our pets are not miniature humans.

The question now arises as to how the perceiver comes to perceive

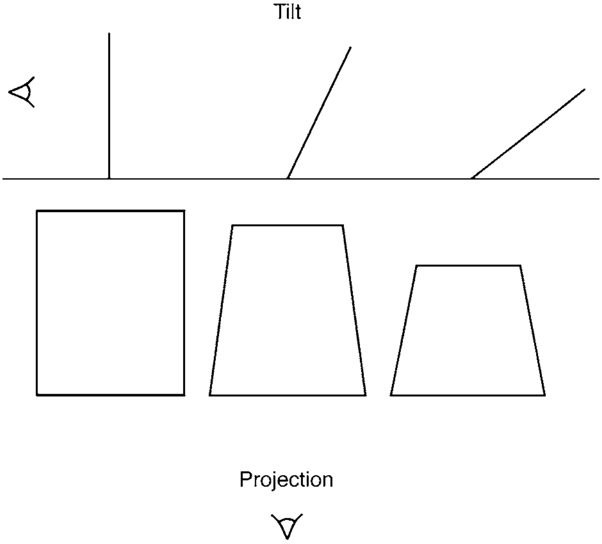


Figure 6.6 The projected shape of a tilted rectangle is trapezoidal.

invariants and affordances. Is it necessary to learn which invariant properties of an array are useful? Are we born able to detect affordances? How do we extract the higher-order information contained in optical arrays? Supporters of direct perception theory have addressed these questions, and we shall now attempt to summarize their main conclusions.

On the question of whether or not perception is learned, Gibson and his followers would remind us that we and other successful organisms are the products of millions of years of evolution. The environmental niches in which our sense organs operate have been responsible for the evolution of the organs; as they have shaped their structure, so also must they have shaped their performance. Thus, learning has indeed occurred in the development of perception, but this has been during the history of the species, rather than the lifetime of the individual. The same sort of evolutionary pressures that have 'taught' our kidneys to respond correctly from birth could have shaped our visual systems to respond in certain ways to contours or gradients of texture.

There is, however, a role for learning during the history of the individual perceiver: humans (and many other 'higher' animals) must surely learn which affordances can be relied upon to satisfy certain goals. That is to say, although the invariant stimulus properties comprising the affordance of, say, 'graspableness' may be perceived immediately, knowing when graspableness is an appropriate property to search for is situationally determined and must presumably be learned. The role of learning in perception is to educate attention.

Resonance

If the perception of the real world involves detecting appropriate invariances in the rich and ever-changing sea of stimulus energy, it is natural to ask how these invariants are detected. Is there, for example, a moment in the visual process when the relative motions of different texture elements are extracted, possibly by correlating their directions and speeds in order to detect texture flow? Gibson's reply to this question would be that it is the wrong way to begin to consider the problem. Such a way of conceptualizing perceptual processes leads to the reductionism, the reliance on hypothetical stages within processes, that disfigures the empiricist approach to perception.

A direct perception theorist knows that there are identifiable peripheral processes to be observed in receptor systems (such as the contour sharpening brought about by lateral inhibition in the retina), but the conclusion would still be that the reductionist approach is wrong. Gibson and his supporters argue that the response to stimulation is a response involving the whole organism. The nature of this response is described by Gibson as follows:

> I suggest that the nervous system operates in circular loops and that information is never conveyed but extracted by the picking up of invariants over time . . . a perceptual system does not respond to stimuli (although a receptor does) but extracts invariants.
>
> (Gibson, 1976)

Once again, the active role of the perceiver in extracting informative, invariant patterns is being stressed. Later, Gibson extended his notion of information pickup by likening perception to a process of *resonance*, which he explains by analogy with a radio set. To elaborate: the space in any room in a modern city is filled with electromagnetic radiation broadcast from large numbers of transmitters, some close at hand, some many miles away. This radiation is non-random: it can convey information. On switching on a radio, all we may hear is the hissing noise arising from its own circuits. But on tuning the radio, we may suddenly hear speech or music: it is now set to resonate with the information available in the electromagnetic radiation. We are witnessing a process of *information pickup*.

The direct perception theorist can now challenge us with this question: 'In which part of the radio is that particular sound being processed?' The answer must be that it is everywhere, for all parts of the radio's circuit are active during the transduction of the radio waves into audible music. Remove any one part of the circuit and the set will fail. But that part cannot then be said to 'compute' music or speech – these are rendered audible by the behaviour of the whole radio, with its components acting together as a single system. The radio is not a perfect analogy, of course, for it is a passive device – we do the tuning. But during perception information is obtained, rather than imposed.

Thinking along such lines suggests that the nervous system may be better modelled by analogue rather than digital devices. To reinforce this point, consider an old-fashioned slide rule. This is a means of multiplying and dividing pairs of numbers without engaging in common arithmetic; one simply adjusts two scales (adding or subtracting two quantities) and reads off the answer. The trick, of course is that the markings on the rule are drawn to represent logarithmic quantities. The necessary mathematics has been built into the structure of the device. Is this the role of evolution in the shaping of the senses?

Realism

It should be clear from this brief introduction to direct perception and ecological optics that Gibson and his followers assume a philosophical position: direct realism. Stated very simply, direct realism is the assumption that there is an external world of objects and that we can become aware of these as a result of our perceptions: proximal stimuli can specify distal ones. This doctrine is contrasted by Gibson with the position that, as our senses must intervene between external objects and our experience of them, all that we can be directly aware of must be sensations or sense-data. That is to say, we cannot experience a hot object directly, but must construct this percept from sensations of heat, touch, and pain.

Gibson accepted that sensations do exist and that we can be made aware of them by training or by adopting certain mental sets. We can be aware of our own physiological states; and no other person can experience that vague presence in our visual field created by our nose.

> Physical acoustics tells the man in the street that sensations of loudness, pitch, and pitch mixture are in his head, and only arise because they correspond to the variables of sound waves in the air. He could not possibly hear a mechanical event; he can only infer it from the data. But nevertheless he goes on hearing natural events like rubbing, scraping, rolling, and brushing, or vocal events like growling, barking, singing, and croaking, or carpenter's events like sawing, pounding, filing, and chopping. Ecological acoustics would tell him that the vibratory event, the source of the waves, is specified in certain invariant properties of the wave train . . . Information about the event is physically present in the air surrounding the event. If the man is within earshot, he hears the event.
>
> (Gibson, 1967a)

This is as clear a statement of his position as we can find in Gibson's writings. Naturally, the philosophical differences between direct realists and their rivals are debated at greater length than this. The arguments can be quite complicated, as may be seen by consulting some of the references to be given later.

An evaluation of the theory of direct perception

We come now to a general assessment of the direct perception tradition that Gibson founded during his career. We shall list what seem to be the most pertinent criticisms to have been levelled against direct perception. Then we shall describe some of the achievements of this new approach.

The meaning of 'direct'

A valuable debate on direct perception (Ullman, 1980) begins with an interesting question. What does it mean, asks Ullman, to say that any process is direct? We shall offer a slightly modified version of Ullman's illuminating analogy. Consider arithmetical multiplication. The input to the 'process' of multiplication is a pair of numbers, the output is their product. Now there is a way of making such a process direct. This would be to create a lookup table. A range of numbers run across the top and down the side of a matrix. Each row/column intersection contains a product of two numbers. Multiplication done this way could be described as direct. And by this we mean that the process of multiplication *cannot be decomposed*.

In practice, however, such a lookup table would quickly become cumbersome and even unusable – imagine looking for the product of column 257 and row 9367. Over the centuries it has been found that a more powerful and flexible way to multiply is to treat each number as composed of units, tens, hundreds, and so on. We then take the digits in a certain order, multiply, record, and carry over when necessary. But this method of doing multiplication cannot be described as 'direct' – it requires the application of different rules at different stages. This in turn means that, as a process, multiplication of large numbers the traditional way is decomposable – it is therefore indirect.

When supporters of the theory of direct perception use the term 'direct', are they using it in the sense defined by Ullman? If so, some evidence can be brought against them. Arithmetical calculation may seem far removed from perceiving. Here is a more psychological example that can be used against direct perception (it was suggested by an article in the Ullman debate, cited above).

It is quite easy to make an outline wire model of Necker's reversible cube. The simple device shown in Figure 6.7 should be painted matt black for the best effect, then viewed at arm's length with one eye closed.[3] Very soon the viewer will experience a reversal of perspective, such that the far sides of the figure suddenly appear closer than the near ones. If now one slowly twists the cube, it will appear to move the wrong way, that is, against the motion of the hand. Then, when the figure reverses back to its correct orientation, twisting results in normal movement.

---

[3] Prosperous readers can generate this illusion by using the wire structure surrounding a champagne cork.

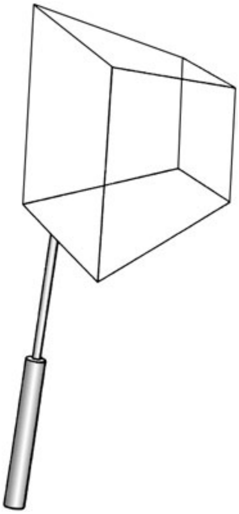


Figure 6.7 A reversible solid. This shape (a truncated pyramid) comprises a large rectangle at the front joined to a smaller rectangle at the rear. When constructed from blackened wire and viewed monocularly, the perspective will occasionally reverse. If the handle is now twisted, the parts of the shape will appear to move in the wrong direction.

The point of this simple demonstration is that although the stimulus array does not change physically, two distinct motions may result from the rotation of the cube. But in order to predict which motion a viewer will experience, it is necessary to know which of the two possible orientations of the cube is being seen. It is quite clear that in this case the perception of orientation (correct or reversed) comes prior to the perception of motion (correct or reversed). Does this not mean that the perception of the cube's motion is decomposable into stages and hence cannot be direct? How would a direct perception theorist respond to this suggestion?

The direct perception theorist might reply as follows. The hollow cube has been *designed* to be difficult to perceive; it has been carefully shaped in such a manner that information in light reflected from it specifies an ambiguous object. And this is true of many famous illusions.

Results from a famous experiment by Shepard and Metzler (1971) can also be used to support Ullman's criticism of the term 'direct'. This was a reaction-time study in which observers had to report, as quickly as possible, whether two shapes flashed onto a screen were or were not identical. The second shape of each pair could in fact be identical to the first, or could be a mirror image of it. The main independent variable in this experiment was the rotational difference between the shapes: for example, both could be identically orientated with respect to the top of the screen, or one shape could be rotated away from the other. The fascinating result was that the times taken to make the key decision – were the shapes identical, or mirror images? – increased linearly with the extent of the rotational difference between them. That is to say, it was as if observers had to engage in a process akin to rotating the representations of shapes before they could make decisions. The greater the necessary rotation, the longer the reaction time.

The role of mental imagery in conscious life has long been argued about by psychologists, some of whom deny that there can be pictures in the head that we can use to solve problems; the problem of representation is a complex one. Nevertheless, the Shepard and Metzler finding does suggest that some process was intervening between the perception of the shapes and the ability to arrive at correct decisions. If it is accepted that this is a reasonable interpretation of Shepard and Metzler's results, then it is likely that this aspect of perception is decomposable and therefore indirect.

A possible answer to Ullman's general question concerning directness and decomposition might be that he has confounded different levels of analysis. Gibson held that perception is direct in the sense that information is extracted directly from optic arrays, and that our awareness of the world is not itself mediated by schemata or representations. He did not deny that mediating processes exist (and he accepted that awareness of the world via words and symbols must be indirect in this sense). But it is the direct relation with *information* that is important – the fact that it is not necessary to decompose it to sensory elements and sensations. However, direct perception theorists have never been very clear about the nature of the processes that mediate this relationship, apart from suggesting that the nervous system somehow 'resonates' with information. Perhaps they should be more explicit about resonance and the sense in which it is a direct process.

The detection of invariants

The detection of invariants is of central importance in the theory of direct perception. Attunement to higher-order patterns within a mass of stimulation forms the basis of awareness. However:

> Although one can criticize certain shortcomings in the quality of Gibson's analysis, its major and, in my view, fatal shortcoming lies at a deeper level and results from a failure to realize two things. First, the detection of physical invariants, like image surfaces, is exactly and precisely an information-processing problem, in modern terminology. And second, he vastly underrated the sheer difficulty of such detection.
>
> (Marr, 1982)

These are important criticisms. Gibson and others believe that there are invariant properties in physical events that afford the perception of those events. But workers in the field of artificial intelligence, such as Marr, have set themselves explicit goals, one of which is to devise systems that will simulate the process of seeing. When these workers try to create some model that will actually extract invariants, they commonly find that it is a very difficult thing to do. It can be argued, of course, that Marr and his colleagues have adopted a faulty model of the perceiver: certainly, the motor activity that Gibson holds to be vital to perception plays a very small part in Marr's theory. But Marr's comments, coming as they do from a distinguished theorist who has tried to simulate seeing, must be taken seriously. This is not to say, of course, that theorists such as Gibson are wrong; rather, that asserting that something must be the case may delude one into thinking that one understands how it is the case. The danger is that a theory may be leading one away from those very problems that it might be fruitful to pursue.

The nature of affordances

The subtlest forms of invariance are affordances. Reading Gibson and others on affordances is rather like reading Freud on dreams: one is convinced at the time, but reflection brings doubt.

It is clear that to know and describe the relatively straightforward invariances used in, say, the perception of space, is a formidable task. But if certain objects in the world directly afford eating, just what is it in the nature of the optic array that can make explicit this affordance? In Gibson's terms the answer (in the visual modality) must be some nested array of solid visual angles; but, as he admitted, we do not really have any idea of the characteristics of such a complex array, and the answer must be many times more complicated than that to the already formidable problem of the spatial invariants.[4] The situation is even worse than this: even if we could define some affordances for a perceiver, it would still be hard to predict behaviour. This is because, in terms of the theory, organisms may have to learn to attend to particular affordances. Before we can predict behaviour, we must know not only the affordances available but also the perceiver's current attentional state. This is a formidable requirement.

On the general question of affordances, however, there are reasons for optimism. We shall attempt to justify this claim in a later section on more modern research. For now, we shall simply remind the reader that the essence of affordances is that they are always *relationships* between organisms and their environments. It follows that to insist upon predictive power is to fall into the animal–environment dualism that direct perception wishes to avoid.

Can affordances ever mislead? Gibson changed his position on this aspect of affordances. Initially, Gibson held that a surface afforded walking, and that was that. But as Costall (1981) points out, '. . . surfaces are not where all the action is'. They may or may not afford walking – ground covered

---

[4] At this point the reader may wish to return briefly to Chapter 3 and our description of techniques used by Purves, Lotto and their colleagues.

with a dusting of snow does, but a thinly frozen lake does not. Later Gibson modified his statements on affordances to recognize their probabilistic nature, thus moving much closer to Brunswik's earlier statements (see Chapter 3) concerning the ecological validity of cues.

Resonance

The idea that a nervous system mediating some form of perception behaves in a holistic manner, resonating to invariant properties among stimuli, is initially attractive, particularly to anyone who has waded through innumerable 'stage' models of perception. On reflection, however, we are forced to conclude that resonance is barely more than an interesting and novel speculation. With stage models of perception we can at least be sure, most of the time, what we could expect to observe in the nervous system, had we the necessary techniques. More importantly for the psychologist, we have hypotheses about the temporal and logical ordering of processes. We should expect, for example, retinal sharpening or filtering to occur before binocular fusion processes; that the recognition of familiar forms would come even later, when meaning has been 'added' to the input; and so on. But how are we to observe resonance? If the answer is, whenever an organism is functioning as though properly in touch with the real world, then this amounts to a tautology. If behaviour was not appropriate or adaptive then we would not wish to invoke resonance.

We can ask whether the process of resonance could *in principle* be observed using new techniques of anatomical observation. Does the nervous system resonate to different modalities simultaneously (it surely must in bimodal perception), and if so, is there a cost to be paid in terms of capacity? Do all nervous systems resonate in their own particular ways, or has evolution produced only one form of resonance? The truth is that we are told very little about resonance. The metaphors used (the radio, the slide rule) are intriguing and stimulating, but they are only metaphors. It could be said that a demand for neurological plausibility is unreasonable at this stage of our knowledge of perception, and is to site the problem at the wrong level of discourse. But then we should be given some guidance as to a possible operational definition of the term 'resonance', or the type of evidence that would convince its proponents that they were wrong.

In response to such objections, a direct perception theorist might reply as follows. A criticism that focuses on the nervous system, asking where resonance occurs, misses the point: resonance is a relationship between the perceiver and the environment. Until there is much more knowledge about the nature of such relationships, and until we can learn to stop thinking about organisms in isolation, it is pointless to look for a place where resonance occurs: resonance is not that sort of concept.

There may now be, however, a better answer to criticisms concerning resonance. As we attempted to show in an earlier chapter, parallel distributed processing models (or connectionist networks) have some striking features that distinguish them from von Neuman machines. Of great relevance here is the fact that what is represented in a network is, in a very important sense, represented all over the network, not in any one special place. Thus, when a network learns to discriminate, say, a male from a female face, the learning is a property of the entire network: the knowledge is everywhere. Is it too fanciful to conclude that such a network is resonating, and in this manner is arriving at the 'answer' to the question implied by the input? The present author feels very optimistic concerning the relevance of such a mechanism for the interesting concept of resonance. Readers may choose to be more sceptical and interpret the above remarks as merely an analogy combined with another analogy.

Direct perception and traditional, laboratory-based research

To end this section we shall simply assert that Gibson and his followers, when writing about the importance of invariants and affordances and the types of research that psychologists should do, have a tendency to underestimate the achievements of the single-variable type of experiment to which they are opposed. This may be for polemical reasons, of course. But is this attitude fair? For example, much of what we know about human perception has come from what were, originally, casual or accidental observations. Any careful observer, at any time in the past, could have noticed the patterning of optical flow and its relationship to our movements and position in space. Of course, it took an intelligent researcher to explore this phenomenon to the point when it could be embedded in a convincing perceptual theory; nevertheless, the phenomenon was there to see, easily controlled and easily manipulated.

In contrast, how do we know that infra-red radiation affords prey detection by snakes? Simply observing snakes in a natural environment won't do: when the prey moves in the dark there are changes in sound and smell, as well as in the direction and strength of infra-red radiation. Only careful experimental studies, in which all variables save infra-red are controlled, can convince us that we have isolated the correct invariant.

Here is another powerful counter-example to criticisms of traditional, laboratory research. It is a fact, long recognized, that when the perception of an object is difficult and its shape and identity elusive, movement of the object (or movement around it) commonly resolves any ambiguity. We have all had numerous experiences of this kind of thing: the brown patch against the tree becomes an owl as we approach; the two-headed monster in the field, we discover, is a pair of cows. And of course, the importance of the observer's movement is repeatedly stressed in Gibson's writings. An interesting question now arises: how many views, how much movement do we need, in order to see the uniqueness of any shape? The answer is contained in a new theorem, unknown until recently even among mathematicians: a shape is uniquely specified by *three views of four non-coplanar points* (Ullman, 1979). This is a very important gain in our understanding of the perception of three-dimensional objects. However, the research leading to this discovery consisted of experiments employing highly simplified stimuli, often displayed under very artificial conditions, such as brief exposures, or the casting of shadows onto screens. The result, however, has been an undoubted success.

Finally, we may comment on a tendency in writings on direct perception to define problems out of existence. Evidence of this has been given when discussing the extraction of invariants and Marr's appraisal of Gibson's work. We can only repeat our earlier point that to say that something simply is the case, may be to lose one's grip on a real problem. In an interesting and provocative section on learning, Michaels and Carello challenge the concept of memory:

> And just as we do not need a vessel in which *ancient* history is brought to bear on the present, we do not need a vessel (memory) in which *recent* history is brought to bear. Plainly and simply, experience changes the animal.
>
> (Michaels & Carello, 1981, italics in original)

If it *is* so plain and simple, why has it been so difficult to discover the laws of learning and forgetting after a century of research?

More recent research

Philosophical issues raised by Gibson's theory

There have been attempts to make philosophical refinements to Gibson's theory since his death. For example, more recent supporters of direct perception have been rightly concerned over the implications of Gibson's use of the term 'realism'. Fodor and Pylyshyn (1988), Noble (1981), Reed (1982, 1987), Katz (1987), and Costall (1981) have all contributed towards a better understanding of the issues involved.

Katz (1987) has examined Gibson's interpretation of the term 'realism'. Had Gibson really adhered to a basic form of realism, Katz argues, then his theory could not have taken the form that it did. To assume the existence of an objective world, independent of perceivers, and also perceivers who are in but separate from that world, leads to some serious problems: 'How could one conceive an ultimate structure that applies in all conceivable circumstances, from every imaginable point of view?' Perception, says Katz, is a matter of circumstances '. . . determined jointly by subject and by object'. If there is only one world to be perceived, how can perception in one species differ from that of another? And how can we explain errors in perceiving? But Gibson constantly stresses the need to consider (for example) affordances in terms both of the world and the perceiver. For this reason, Katz suggests that Gibson's is really a *relativist* rather than a naive realist.

Costall (1981) is concerned with the same issue. As organisms play an active role in the creation of environments, we must abandon Gibson's distinction between the objective world and perceivers that is implied by statements on realism. In other words, when Gibson discusses realism, he tends, like others, to treat the organism as nothing but a perceiver – a view that his general work was aimed at denying. Costall cites modern biologists who also reject the idea of the environment as a 'pre-existing slot' within which the organism must fit. Costall makes a strong case for what he describes as *mutualism*. And if we acknowledge that the world has changed since the beginning of life and that organisms indirectly influence their environment, we must concede that '. . . in an important sense, the world is other organisms'. Costall hopes that a stress on mutualism will provide a much sounder underpinning for the framework of direct perception.

Noble (1981) has made a detailed study of the origins of the indirect approach to awareness. He traces it back to Descartes's corporeal ideas hypothesis. In essence, Descartes argued that, as sensations do not resemble the objects that cause them (as in tickling, for example), there must be two distinct worlds: the world of objects, and the world of thinking creatures. What follows from this is the concept of mental processes operating on the 'deliverances' of the senses. This is, as we have shown, the basic form of an argument for indirect awareness and perception. Noble's contribution is to show what a long history this idea has had, and how it has been refined over the years until it has become interwoven into the fabric of psychological thought. What was originally a scientific hypothesis has become dogma, and this in turn is the source of some of the resistance to the new paradigm represented by direct perception.

The papers cited in this section should be consulted by anyone who wishes to learn more about the effort that has been put into the philosophical refinement of the theory of direct perception.

We turn next to an outline of a small sample of the ingenious experiments to have been inspired by Gibson's ideas. At the very least, they may demonstrate how fruitful his ideas continue to be.

Movement and vision

Gibson argued against the distinction between sensory and motor systems and stressed the importance of movement in visual perception. An interesting analysis by Coren (1986) lends indirect support to Gibson's position.

Coren examined the phylum Mollusca. This phylum of animals contains 50,000 known species, making it second in size only to the arthropods in the animal kingdom. Coren points out that molluscs vary greatly in the degree to which they move. At one extreme are the bivalves, which generally move very little in their habitats. At the other extreme are the cephalopods – including octopus, squid, nautilus and cuttlefish – many of which can move very rapidly and nimbly. If we compare the visual systems of these two groups, striking differences emerge. The bivalves have rows of primitive eyes that are capable only of rather crude visual discriminations. This is easily explained in evolutionary terms: if an animal lives attached to rocks at the bottom of murky sea water, there will be no evolutionary pressure to develop sophisticated eyes. In contrast, creatures such as the octopus and squid have very well developed eyes and can make fine visual discriminations. Anyone who reads Coren's paper will find themselves agreeing with Gibson's argument.

Affordances

Accounts of research on affordances by post-Gibsonian psychologists tend to use certain special terms and concepts. It may help readers who wish to pursue this topic further and read the experimental literature to begin with a short exposition of some relevant terms before describing some of this newer research.

Extrinsic measures are objective measures of some aspect of an object or situation expressed in standard units such as grams, metres, and seconds. *Intrinsic measures* are extrinsic measures that have been re-scaled in terms of some dimension of the observer or actor in a situation. Some intrinsic measures are defined as *pi numbers*, which are dimensionless, body-scaled ratios, useful when describing the fit between an organism and its environment. As the fit between an organism and its environment is altered, so is the nature of the affordance (and the value of pi). The optimum point of such a fit corresponds to a 'best' match between the organism and the environment; it will be the preferred value of the affordance and will be associated with stable, maximally efficient behaviour. Further changes in the fit will produce a critical point corresponding to a phase transition or critical boundary in behaviour. Thus, for any human there will be a walking speed that is most efficient and most comfortable. As readers will be aware, speeding up one's walking becomes increasingly uncomfortable until, abruptly, one breaks into a run. That is to say, slow running is more comfortable (and more efficient physiologically) than very fast walking. The change from walking (where part of the body is always in contact with the ground) to running (where the body leaves the ground between steps) can be described as a change of *gait*.[5]

We shall now give a detailed account of one of the most interesting attempts to find and measure an affordance.

A study of stair climbing (Warren, 1984)

Gibson recommended the study of the environment in which organisms evolved. It may seem a strange leap from this to the study of people's

---

[5] Some animals have a variety of gaits. Horses walk, trot, canter, and gallop. Cats have gaits, but the present author cannot decide how many.

perception of staircases. However, we can think of stairs as simplified versions of uneven terrain and hope that it will become possible in the near future to extend the techniques developed in the study to be described to more natural surfaces.

The Warren experiment has been selected for two reasons: first, its thoroughness – it is hard to see how it could have been better designed and carried out; second, the surprisingly clear-cut and fascinating findings that emerged.

The stair-climbing variables selected by Warren were riser height (R), the height of each vertical step, which combines with tread depth (T) to give the stair diagonal (D). The person climbing the stairs has mass, climbs at some favoured rate, and is limited in the size of vertical step he or she can take by the overall leg length (L), thigh length (L_1) and lower leg length (L_2).

As Warren points out in his introduction, the search for a formula for ideal staircases has a long history. Clearly, very shallow staircases (low riser height, deep tread) are inefficient: it takes too long to get high. On the other hand, one can easily think of staircases with riser heights so great that few could climb them. What architects have sought is some compromise that fits the average person and allows most efficient climbing. The French architect Blondel (1675–1683) suggested that as a comfortable pace length was 24 (French royal) inches, 2 in should be subtracted from tread depth (T) for every inch of riser height (R):

$$2R + T = 24 \text{ in} \tag{1}$$

A similar formula was still in use as late as 1978.

Given that an affordance is a *relationship* between an organism and the environment, Warren's aim was to investigate the relationships between staircases and climbers of the staircases, and then to express these as dimensionless pi numbers. If pi numbers can successfully predict behaviour and changes in behaviour (critical points), then a new affordance would have been discovered. Warren began by focusing on the relationship between riser height and leg length. The expression linking these two can be expressed as a pi number:

$$\text{pi} = R/L \tag{2}$$

where R = riser height and L = leg length.

Figure 6.8 shows a mechanical model of a climber. Clearly, the maximum riser height that a person can use will demand maximum leg flexion; any greater height will require the person to jump or make use of the hands.

The lengths of the legs (L = total leg length) and lower legs of groups of tall and short observers were measured. From the model shown in Figure 6.8, it was calculated that critical riser height must be:

$$R_c = L + L_1 - L_2 \tag{3}$$

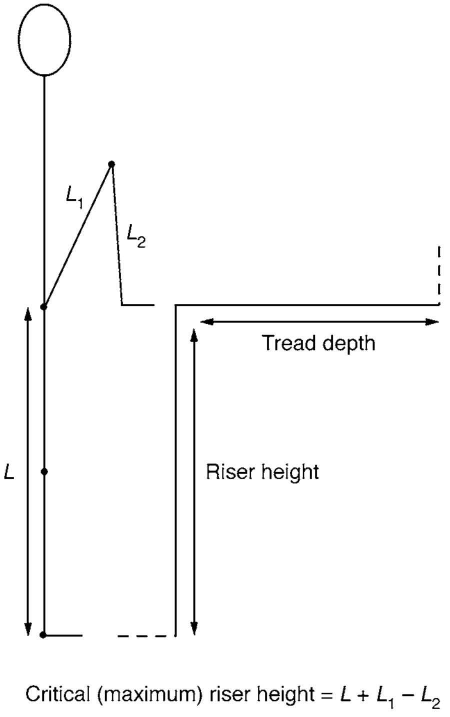


Figure 6.8 The anatomical limits to stair climbing. Note that although the raised thigh adds to the height of the raised foot, the lower leg subtracts from it. This sets a limit on the height of a step that can be mounted using the legs alone.

where L = total leg length, L_1 = length of the thigh, L_2 = length of the lower leg, i.e., total leg length plus the length of the flexed thigh (which is adding to the height the foot can attain), minus the length of the lower leg (which hangs downward, therefore subtracting from the maximum height the foot can attain). When anatomical data from the tall and short groups of observers were substituted into the above formula, the result was a pi value of 0.88 L for both. Note that as the ratio of lower and upper leg segments tends to be constant in humans, this value of pi also tends to be constant across people of different heights.

The two groups of observers were then shown a series of projected life-size pictures of an experimental staircase. This was built in such a manner that the riser heights could be changed. The various versions of the staircase had been photographed from two different vertical positions corresponding to the mean eye heights of the tall and short observers, respectively. The observers then rated each staircase on a seven-point scale, indicating whether they considered it to be climbable or not. An analysis of these ratings showed that the percentage of 'climbable' judgements dropped from 100% to 0% as riser height was increased. The riser height at which mean ratings of each group crossed the chance or 50% value was taken as a perceptual boundary for that group. The outcome was that the *perceived* critical riser height for the tall group was 0.88 *L*, that for the short group 0.89 *L*. Compare these with the value of pi calculated from anthropometric data (0.88). This is an impressive result, but there is better to come.

Three short and three tall male volunteer observers climbed an adjustable moving staircase. As they did so, gas samples of their breathing were collected and oxygen and carbon dioxide analyses performed. From these it was possible to calculate the rate of energy expenditure required to mount different riser heights at various climbing rates.

The search now was for *optimum* rather than critical values of pi. Multiple regression analyses, followed by a set of curve-fitting exercises, permitted the calculation (for tall and short groups of observers) of that riser height associated with maximum ascent at minimum energy cost. The optimum value of pi was found to be the same for both groups of observers. With optimum riser height = R_o, and total leg length = L, then it was found that for both groups of observers:

$$\text{pi} = R_o/L = 0.26 \tag{4}$$

Following this finding, is it possible to ask whether observers are capable of visually detecting the optimum riser height to suit them?

Once again, tall and short groups of observers were asked to judge a series of black and white pictures of stairways having six different riser heights. However, on this occasion they were asked to judge the relative ease of use of the staircases. The rest of the experiment was essentially the same as the first one. The results were striking. For both groups of observers (using the same terms as above), pi = $R_o/L = 0.25$. There is thus an extraordinarily close match between an optimally efficient riser height and an observer's visual perception of that height: the affordance is detectable.

In summary, Warren's results tell us that the *critical* riser height for an individual (the point at which his or her behaviour must change from climbing to jumping or using hands and knees) is equal to 0.88 of his or her total leg length. To find that riser height which affords *optimum* stair climbing, multiply the climber's leg length by 0.25.

There is one very interesting aspect of the result for the visually preferred riser height (the optimal affordance). The bodily dimensions of Americans are well known as a result of large-scale anthropometric surveys. Using Warren's pi number in combination with data on the distribution of leg lengths in the American population yields a value for optimum riser height that is considerably higher than the riser heights commonly used in stairways. Thus, Warren's observers did not simply express preferences for the familiar. His results do not therefore seem to reflect learning processes. This experiment could serve as a model for others wishing to find quantifiable data concerning affordances.

Infant perception studies

These were described at some length in Chaper 2. At this point we should point out that Eleanor Gibson, Gibson's partner until his death, is also a distinguished experimental psychologist. She has reviewed the concept of the affordance from a developmental point of view (E. Gibson, 1982). In a convincing argument for the importance of the affordance for theories of development, she takes as a starting point the claim that:

> Perceiving an affordance implies perception that is meaningful, unitary, utilitarian, and continuous over time to the extent that environmental events that pertain to the observer may require. To what extent must young creatures (human or otherwise) learn to perceive them? And if they must learn, how is it done?
>
> (E. Gibson, 1982)

Gibson then reviews some of the research that is relevant to these questions. Work on 'graspability', for example, appears to show that objects of graspable size are responded to differently from non-graspable objects by the age of 3 months. This discrimination is revealed by patterns of hand and arm movements towards the objects. Other studies have shown that when infants aged about 14 weeks put their hands out to grasp moving objects, they move them to positions *where the objects will be*: they seem able to extract the relevant affordance for prediction. When 3-month-old infants are habituated to the sight of objects that have been subjected to certain rigid transformations (rotations around horizontal and vertical axes, for example), a non-rigid deformation (squeezing) causes the object to be attended to once more. Infants thus appear capable of distinguishing between these two fundamental ways in which objects can change.

It has been found that by 6 weeks infants will blink when faced with a looming object, that they are sensitive to optical information concerning impending collision. (Parents reading this should sit down again – these infants never actually get bumped.)

Another reviewer, von Hofsten (1983), describes experiments that have shown that infants can fuse information across modalities by an early age. For example, when viewing two moving films (shown simultaneously, side-by-side) of an object rising and falling, they prefer (they spend a longer time looking at) that film that is synchronized with the appropriate sound of a contact with the ground.

Von Hofsten reasons that the stability of our perceived world is vitally dependent upon our ability to perceive the permanence of objects during changes across space and time. Studies have shown that infants at 8 months will attempt to retrieve hidden objects, and that when an object goes behind an occluding surface and a different one emerges on the other side, the infants show signs of surprise.

Butterworth's (1983, 1988) detailed reviews include many infant studies that are directly relevant to the theory of direct perception and should be consulted in full by interested readers (some of this work has already been described in Chapter 2). In another study reviewed by Butterworth (Granrud et al., 1984), infants watched computer-generated displays of randomly moving dots. An impression of discontinuity at an edge was created by having part of the texture on the screen continuously deleted by the remaining texture. To adult observers, this looked like one moving surface sliding behind another (such occluding effects are referred to frequently by Gibson in his later writings). It is known that infants, faced with a choice, will attempt to grasp the nearer of two objects. In this study, infants aged 5 months reached and attempted to touch the television screen at a position where one surface appeared to be above the other. It seems, therefore, as though infants can use dynamic properties of stimulation to acquire knowledge of depth.

His study of published research in this area leads Butterworth to suggest that the vital distinction that each of us must acquire – that between ourselves and the world – is imposed very early in life by the structure of the optic array, which is what Gibson would have predicted.

As was said at the start of this section, material similar to that above was discussed in Chapter 2 on the Gestalt theory. It has been included here also because, for the theorist, this area of psychology is potentially very important. It may well be that results from infant studies will eventually help us to decide upon the correctness (or otherwise) of key parts of the theory of direct perception. It can be said here that the ability of infants to respond to higher-order, invariant properties of stimulation is looking more and more impressive. Gibson may have been correct in believing that such abilities have been acquired through the course of evolution and do not have to be learned during the development of the individual. At the very least, it can be claimed that the perception of some invariances comes very easily and naturally to the human infant: a fact that supports direct perception but which will have to be accounted for by any general theory.

A new invariant: the cardioidal strain transformation

When we see a familiar person who has aged, our perception of the face tells us two things: first, that the face has changed; second, that it is, however, the same face. This is a complex situation involving simultaneous perception of identity and change. What is the basis of this physiognomic perception? How can we see the continuing identity of a face? A partial answer is that some of the important changes that occur during ageing can be described by a mathematical function, the cardioidal strain transformation

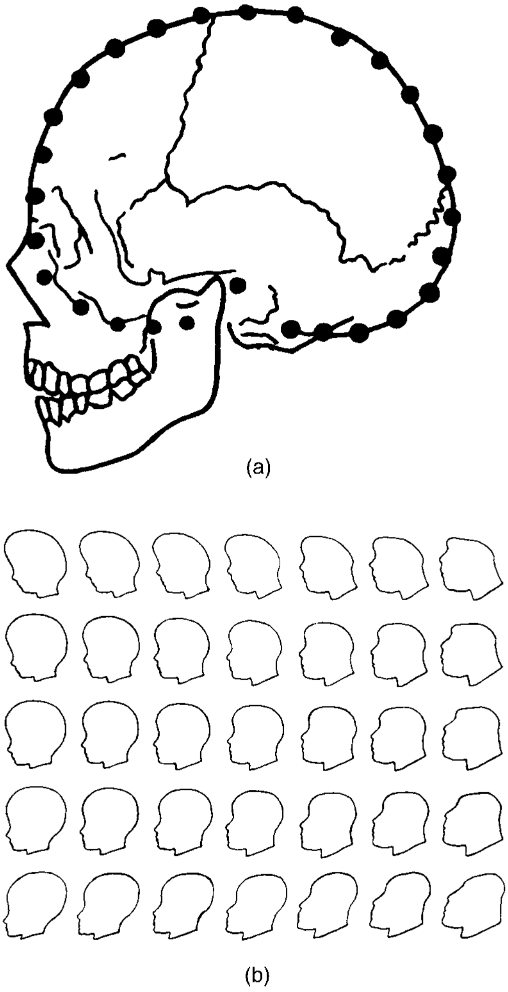


Figure 6.9 (a) A cardioid fitted to the outline of a skull (reprinted from *Perception: Essays in Honor of James J. Gibson*, edited by R. MacLeod & H. Pick Jr. Copyright © 1974 by Cornell University. Used by permission of the publisher, Cornell University Press); (b) outline drawings of a face made to age in appearance by submitting the original outline to a series of cardioidal strains (from Pittenger & Shaw, 1975. Reproduced with permission from the Psychonomic Society).

(see Figure 6.9). The outline of a skull can be fitted by a cardioid. As the skull ages, small changes in the parameters of the cardioid can match the changing shape. And it is changes in the shape of the skull that are partially responsible for changes in our faces as we grow up. Now when a sketch of a face is subjected to controlled distortion by cardioidal strain, it appears to age (Pittenger & Shaw, 1975). It seems clear that there is something that persists during the ageing of a face, an invariant that can be recovered from the cardioidal function, and which we seem to be able to perceive. What seemed like a very difficult problem has begun to yield to the application of the concept of the invariant.

Studies of optic flow

The first study to be described under this heading is that by Lee and Reddish (1981). Researchers such as these are beginning to understand optic flow, the importance of which Gibson stressed repeatedly in his writings.

When the gannet (*Sula bassana*) sights fish, it dives at the surface of the water, often from heights of up to 30 metres. During the dive, the gannet adopts an increasingly swept-back wing posture. During the dive the wings, although swept back, are not fully so: the partial closure leaves enough wing for steering. However, the terminal velocity attained during a dive may reach 50 mph: if the gannet entered the water at this speed with its wings partially deployed it would damage itself. Then, immediately prior to hitting the surface, the bird stretches its wings straight back to enter the water in the most streamlined posture it can adopt (see Figure 6.10). What information is available to the gannet to allow it to time its motor behaviour so precisely? Lee et al. have proved that there is a regularity in the changing visual image of the textured surface of the sea. This constant is the inverse of the rate of dilation of the visual image of texture elements. Lee and Reddish named this constant *tau*. They then constructed a model, that includes tau,

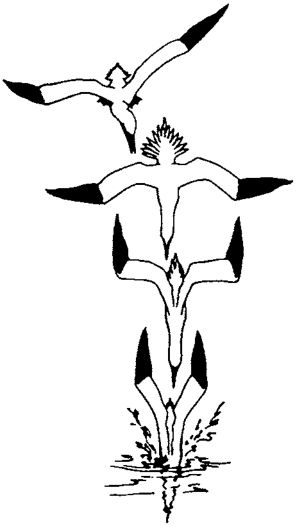


Figure 6.10 Changing wing positions of a diving gannet, *Sula bassana*. How does the gannet time these changes? From Nelson (2000). Reproduced with permission of the originator, Fenix Books.

to show how information specifying time-to-contact can be extracted during a dive. Then they tested data from films of actual diving behaviour against their model. The fit to a curve based on their equation is a good one.

Lee's interest in optic flow and related phenomena has continued. Together with his colleagues, Lee has investigated aerial docking by hummingbirds (Lee, Reddish, & Rand, 1991), the ability of somersaulters to land on their feet (Lee, Young, & Rewt, 1992), and how golfers guide their swings when putting (Craig, Delay, Grealy, & Lee, 2000). This is a fascinating combination of practical and theoretical problems and shows that the tradition founded by Gibson is in excellent health.

The second study to be described, Dienes and McLeod (1993), is not directly about optic flow phenomena as such, but is so close in spirit to some of Lee's researches outlined above that it can be included in this section.

How does one catch a ball? Dienes and McLeod actually refer to a cricket ball, but their answer is a general one (and is quite fascinating). The ball is hit and flies upwards and towards the boundary. If a fielder is to catch the ball, it is clearly necessary to be in the right place as the ball nears the ground. But how far and how fast to run? The answer (arrived at by geometric reasoning and empirical research) is wonderfully simple. The ball rises and then falls in a parabola. Let us assume that the catcher makes a decision to go for the catch when the ball first starts to descend. If a = the angle of the catcher's gaze to the ball, then he or she must run at a speed such that:

$$d^2(\tan a)/dt^2 = 0 \qquad (5)$$

This is to say that the catcher must keep the acceleration of the tangent of the angle of gaze with respect to the horizontal at zero. The acceleration of the tangent is equivalent to the acceleration of the vertical projection of the ball. When this value is negative, the catcher is running too slowly and the ball will drop to the ground in front of him or her; when positive, the catcher will overshoot the point where the ball will land. So catching may require changes in running speed. When running speed is correct, and equation (5) is satisfied, the catcher's gaze (and head position) will also be changing at the correct rate.[6] Readers should now go out and attempt to catch some high balls in an attempt to appreciate this convincing account of the perceptual basis of a familiar skill.

In a very recent study, Zaal and Michaels (2003) have extended the type of work described immediately above by using a completely new technique: an automated virtual environment. This comprises a room with three walls and a floor within which an observer can move freely. Computer-generated

---

[6] Dienes and McLeod (1993) state that changes in head position could be monitored by the vestibular system of the inner ear, or could be based on unconscious calculations using the perceived distance and height of the ball.

images can be projected into the room, and in this experiment the images could include moving 'virtual balls'. In one condition the observer is instructed to try to intercept the balls. It was found that performance with the hand was less good than that when the forehead was used in the interception. So far, this type of research has not yet discovered anything dramatic about visual perception. However, now that it is possible to create virtual environments in this manner, we may expect rapid progress to be made in exploring the characteristics of optic flow as an invariant. Once again, we note the potential influence of technological developments on research and theory – a claim that has been made several times before in this volume.

To end this sub-section, we must stress that the study of optic flow becomes much more complicated when the movement of an observer towards a surface or object is not straight on. In fact, even in the more straightforward condition of straight ahead condition, Gibson's mathematical analysis was wrong. In situations when, for example, we are walking or driving, it is commonly necessary to estimate the chances of collision with an object or surface towards which we are moving *obliquely*. Similarly, we don't always look straight ahead when moving. The resulting oblique relationships between angle of gaze, direction of movement, and converging paths are much more difficult to analyse. Interested readers should read Cutting (1982, 1986; Cutting et al., 1995), who has made major contributions to the study of perception during motion by enriching our knowledge of the complexities of optic flow.

Affordances and robotics

The present author was quite surprised when he came across this subject during a literature search. What link can there possibly be between Gibson's concept of the affordances and the design of robots? When reading some of the relevant literature on robotics, the link became quite clear. Affordances, as defined by Gibson, are about grasping things, avoiding obstacles, climbing over objects, crossing surfaces, and so on. Now if one is going to build a machine capable of moving around in the world, it will have to grasp things, move over surfaces, avoid obstacles, and so on. It is for this reason that a number of workers in the field of robotics have profited from a study of Gibson's work. The present author knows too little about robotics to assess the quality of this work, except to appreciate the final performance of any robot. In order to give more competent readers a possible start when searching the relevant literature, we refer to just a single reference. This is a paper by Duchon, Warren, and Kaelbling (1995) (incidentally, Warren is the author whose work on stair climbing was described earlier in this chapter).

Direct perception and design

When, some years ago, the present author began to describe in his lectures some of the work described below, the students became extremely enthusiastic: so much so that the author was persuaded to create a new Third-Year seminar on Psychology and Design. The reason for this enthusiasm is not hard to understand (and has nothing to do with the author's ability as a teacher). When young people enrol on a university course in Psychology, they have hopes that this will teach them something about the real world. In fact, before one can call oneself a psychologist there is a mass of technical and theoretical material to master. Much of this comes from experiments performed in laboratory researches or from large-scale studies of topics such as IQ or personality differences. This is not the day-to-day world that students inhabit. Offered the chance to examine this world from the viewpoint of a perceptual theorist, but with practical real-world problems in mind, students grasped at this opportunity.

We turn now to a brief description of work that students were so enthusiastic about. The work attempts to use general psychological principles in order to create better designs of objects and systems. The relevance for this chapter is the fact that the work to be described takes as its starting point Gibson's concept of the affordance.

D. A. Norman

Donald Norman is well known for his distinguished contribution to the study of human memory and cognition. He is also someone who became increasingly exasperated by aspects of modern design, whether of objects or of systems. His more popular writings (Norman, 1988, 1992, 1993) contain wonderfully entertaining accounts of his personal battles with computers, doors, refrigerators, and airline schedules. A more technical account of his work is included in Norman and Draper (1986).

> I have studied people making errors – sometimes serious ones – with mechanical devices, light switches and fuses, even airplanes and nuclear power plants. Invariably people feel guilty and either try to hide the error or blame themselves for 'stupidity' or 'clumsiness'. I often have difficulty getting permission to watch: nobody likes to be observed performing badly. I point out that the design is faulty and that others make the same errors. Still, if the task *appears* simple or trivial, then people blame themselves.
>
> (Norman, 1988)

Norman decided to do something about this.

The essence of Norman's position is that humans naturally do some things well, others badly. Bad design fails to recognize this fact. Here are a few examples, drawn from Norman's own writings.

If each key on a telephone has one and only one function, people will quickly learn to use it effortlessly and without error – it will be a pleasure to use. Where the telephone has more functions than keys, its use will require thought and recourse to memory (or the instruction book) – phoning will be less natural, and hence less pleasant. Similarly, we have all struggled to use a strange gas cooker in which the burners are arranged in a square but the controls are in line: which control is linked to which burner? It is necessary to memorize the particular arrangement used (or to peer at the symbols printed on each control knob) in order to select the correct burner. But if the controls are themselves arranged as a square, then the mapping between them and the burners is a natural one and control of the cooker becomes effortless.

We know more than this about errors. Humans profit from feedback to confirm that their actions have been appropriate, and yet in many computer networks one can wait for minutes on end before the (busy) system responds. Having given a computer an erroneous command, the outcome should not be fatal – we should be offered a chance to retrieve the situation.

Basing his analyses on what is known about humans, and using the concept of the affordance, Norman is able to describe many examples of good and bad design:

> Affordances provide strong clues to the operations of things. Plates [on doors] are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction is required. Complex things may require explanation, but simple things should not.
>
> (Norman, 1988)

Norman's writings are an argument for 'user-centred' design. They urge us to oppose the worst that modern technology can do to us, and to insist that this technology enhances rather than diminishes our lives. Norman's writings suggest ways in which this goal could be achieved, and one important way is to make use of affordances. One suspects that Gibson would have been happy to learn of this application of part of his theory.

An attempt at a compromise between empiricism/constructivism and the theory of direct perception

Quite recently, there has been an impressive attempt to reconcile Gibson's position regarding direct perception with the older, constructivist, approach outlined in Chapter 5. This is contained in an article by Joel Norman of the University of Haifa, published in the journal, *Behavioral and Brain Sciences* (Norman, 2002). For those readers who are not familiar with this journal, it can be described as follows: typically, a theorist sends a lengthy article to the journal. Copies of this are then sent to various experts in the field, who write their own reactions to the article. These are published, together with a final summary from the original author, summarizing any criticisms and attempting to deal with them. This is an excellent way of doing things.

Having reviewed the controversy regarding the nature of perception – whether Gibson's ideas or the opposing constructivist views are correct – Norman describes a number of brain studies that, he concludes, can help resolve this conflict.

Two visual systems

A large number of studies have shown that the vertebrate visual cortex contains at least two different systems. This evidence comes from studies of animals that have had part of the visual cortex severed or ablated, studies of the behaviour of patients who have suffered strokes, and studies of people undergoing fMRI scans (as described in Chapter 4).

Norman provides a comprehensive review of these studies in the introduction to his article. The two systems uncovered by researchers are known as: (a) the *dorsal system*, in which the analysis of incoming data is done, with the aim of controlling visually guided behaviour vis-à-vis the environment and the objects within it; (b) in contrast, the *ventral system* is responsible for the recognition and identification of the visual input. The systems differ in other ways: the dorsal system is sensitive to high temporal frequencies, making it better at detecting motion; the ventral system is more sensitive to high spatial frequencies, making it superior in detecting fine details in the stimulus input. In fact, things are a little more complicated than this, and there is a degree of overlap between the two systems. Nevertheless, there is now enough evidence to support the claim that the two systems are, in an important sense, functionally different. This claim is supported by a study by Ungerleider and Mishkin (1982). These researchers found evidence that the two functionally different pathways in the monkey visual cortex were anatomically distinct. The dorsal pathway led from the primary visual cortical area to the posterior parietal cortex; the ventral pathway led to the inferior temporal cortex. It has long been known that the parietal and temporal cortical lobes carry out very different functions, and this was confirmed when Ungerleider and Mishkin created lesions in the two relevant cortical lobes. When the posterior parietal lobe (the terminus of the dorsal pathway) was damaged, the monkeys lost the ability to discriminate between different landmarks; when the inferior temporal region (involving the ventral pathway) was damaged, the monkeys could no longer discriminate between different objects.

Norman reviews a large number of studies that tend to confirm the existence of the two pathways and which explore the functional differences between them. His summary of this evidence leads him to conclude that, basically, while both pathways analyse visual inputs, the main role of the ventral system is the recognition and identification of the visual input. In other words, the input must in some way be compared with stored representations. In contrast, the role of the dorsal system is to analyse the visual input in order to allow visually guided behaviour directed at the environment and objects within it: for example, pointing, reaching, grasping, walking towards, climbing, and so on.

The reader may have anticipated where this story is leading. Might it be, suggests Norman, that the ventral system performs those functions that supporters of the empiricist/constructivist position claim to be the essence of visual perception, while the dorsal system performs those functions that Gibson and his followers hold to be of central importance in seeing the world?

This is a remarkable attempt to reconcile the theories we have outlined in this chapter and in the preceding one. Readers who wish to know more about Norman's argument and the accompanying comments on it should now read his work in full. It is just possible that, had he known of the findings summarized above, Gibson might have accepted a compromise position regarding his theory and the constructivist theory he opposed so vigorously in his lifetime.

Final remarks on direct perception

There is more material in the theoretical writings of Gibson and others than can possibly be summarized in a single chapter. Readers are urged to read some of the references given in this chapter and its endnotes for a more detailed account of a major theory. However, it is our opinion that statements of the theory (once the major points have been grasped) are less interesting in the abstract than when they are coupled to experimental researches and demonstrations. In a sense, this theory of perception is under-specified. Things become more exciting when parts of the theory are tested. This is the reason why this chapter has included so many accounts of experiments. The work on cardioidal strain and ageing faces is more compelling than simple assertions about invariants; the richness (and complexity) of optic flow phenomena became obvious only when researchers attempted to measure them. Solid horizontal surfaces afford walking on – it is hard to disagree with that statement – but how interesting that infants quickly come to perceive that affordance. And now we have a new way of thinking about the things and systems we must deal with in our daily lives: do they offer good affordances?

At the core of the empiricist or constructivist theory of perception is the belief that proximal stimuli cannot fully represent distal objects, and therefore something must be added to incoming information in order to achieve valid perception of the world. The essence of the theory outlined in this chapter is that under 'natural' conditions, that is, in the unbuilt environment, there is a richness and structure in the countless stimuli available to an observer at each moment, such that the world can in fact be specified. Although we must acknowledge the criticisms of Ullman (1980) and others concerning the decomposable (and therefore indirect) nature of perception under some conditions, we should remember that Gibson did not believe that his theory applied to the perception of cultural artefacts.

As Gibson and others have reminded us, the human is just one out of a vast range of perceiving animals. If the theory of direct perception stimulates more research into the perceptual abilities of non-human species, it will have rendered an important service. Phrases such as 'perceptions are hypotheses', convincing at the human level, do not seem to carry as much weight when we look at the behaviour of dragonflies or tapeworms.

This ends our discussion of the work inspired by Gibson's ecological optics and the theory of direct perception. Although we have pointed to what seem to be weaknesses in this work, we have also tried to convey something of the excitement of this way of thinking about perception, and we have outlined the way to a possible compromise between Gibson's theory and its main rival. It is likely that the debate over the correctness or otherwise of Gibson's approach will continue for some time to come. It should be an interesting debate.

Endnotes

- MacLeod and Pick (1974) is a collection of essays written in honour of Gibson.

- Gibson maintained that illusions occur only in artificial conditions. In this he was seriously wrong. Sitting in a stationary train, most of us have experience of smooth movement when a train next to ours starts to move. The horizontal–vertical illusion can be seen in a great many natural situations. In a series of experiments involving solid objects, Day, whose work has been cited in Chapter 5, has found effects analogous to the geometric illusions. These findings detract from the power of Gibson's theory.

- Ullman (1979) opened a debate on direct perception. His critical analysis of direct perception is followed by a series of comments by various researchers arguing for and against the theory. This is an invaluable debate.

- Fodor and Pylyshyn (1981) is a rigorous analysis of the implications of Gibson's theory.

- Gibson always stressed the importance of movement in visual perception and, as we have shown, wanted to remove the distinction between motor and sensory aspects of seeing. Support for this position comes from a study by Coren (1986).

- Gibson was interested in extending his theory to the perception of art. The present author is sceptical concerning the success of this venture. However, readers who wish to judge for themselves should read Gibson (1971b).

- Readers wishing to follow current work on direct perception should consult the journals cited in the final sections of this chapter.

- *The Psychology of Everyday Things* (Norman, 1988) is an interesting and entertaining introduction to D. A. Norman's work on design and his use of the concept of the affordance.

- The best way to gain some impression of modern work using Gibson's ideas in robotics is to use a powerful search engine, such as Google. Enter 'Gibson robotics', or 'ecological optics robotics' and one is soon on one's way.

7 Marr's computational approach to visual perception

This chapter will attempt to outline what many consider to have been the most important developments in the history of perceptual theory. To date the emphasis has been on visual perception, although there are good reasons to believe that successful applications will be made in other sensory modalities. We shall illustrate the approach by describing the work of one man, David Marr, whose contribution can be placed within the context of the new discipline of artificial intelligence.

Marr and others in his discipline concentrate heavily upon processes in the peripheral visual system and show how these extract information from proximal stimuli.[1] At the same time, however, knowledge of the external world is used in a highly original way to impose constraints upon models.

This chapter will cover the following topics:

- David Marr.
- The background to the artificial intelligence (AI) approach.
- Marr's theory of vision and his programme for research.
- Applying the theory.
- Further aspects of Marr's work.
- More recent researches.
- An appraisal of the computational approach to vision.

David Marr

David Marr's first interest was in mathematics which he studied at Cambridge, becoming a Wrangler. After this distinguished start to his career, Marr did graduate work in the department of physiology at Cambridge, where he developed a model of the functioning of the cerebellum. He then learned the techniques of computer modelling at the Massachusetts Institute

---

[1] Once again, we remind readers that some of the material in this chapter will be easier to understand if they can acquire some outline knowledge of the general organization of the visual pathway, from retinal cells to the visual cortex.

of Technology, where he spent the last years of his professional life. Marr had specialized knowledge of mathematics, physiology, computer science, and experimental psychology. His work on artificial intelligence led to numerous papers on perception and, finally, to his book, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, which was published posthumously (Marr, 1982). David Marr died of leukaemia in 1980 at the age of 35.

The background to the artificial intelligence (AI) approach

Forerunners of AI

AI research in general, and the computational approach to vision in particular, are part of an important scientific movement. The historical background to this movement can be outlined by describing three important developments: information theory, cybernetics, and the construction of large digital computers. These developments may seem somewhat irrelevant to the study of visual perception, but in fact they comprise the theoretical tradition out of which the computational approach crystallized.

Information theory was outlined in Chapter 2. As developed by Shannon (1948), the theory made it possible to quantify the information flowing through any system, whether that system was a telephone cable, a television channel, or a person reading a page of text: the measure was essentially neutral with regard to the content of a message. One obvious application of the new calculus was in neurophysiology: a nerve fibre fires according to an all-or-none principle and at certain rates. This transmission of discrete impulses can be viewed as a code, and the rate at which information can be transmitted by one or many neurons may be assessed. Similarly, when a person reacts at maximum speed to one of several possible signals, information theory can be used to assess that person's information-handling capacity. To be able to compare such apparently different situations using an objective measure of information seemed to many to be a very useful development.

Cybernetics, which was developed initially by the mathematician Norbert Wiener, is the application of mathematics to various systems, particularly those that show self-regulation. Initially applied to self-regulating machines, certain concepts from cybernetics quickly proved useful in psychology and physiology. A notable example is 'feedback', which describes how part of the output from a machine can be used to regulate and control the input. 'Negative feedback' typically promotes stability by using the difference between a desired level of output and the actual output level to reduce the input to the system; this damping effect is used to maintain homeostasis in living organisms. 'Positive feedback' tends to have the opposite effect, using output to increase gain and drive the system into instability. Thus, an after-image (which will form if one stares at a bright light) which is off-centre in the visual field will induce reflex pursuit movements of the eyes which try, vainly, to centre the image. The movements cause the image to appear to move even further to one side, and so the pursuit continues, the apparent speed of the after-image getting faster and faster. We remind the reader of an earlier discussion of the impact of digital computers on modern thought in Chapter 5. By the 1950s there were large numbers of these remarkable machines. For many psychologists, they became irresistible as a metaphor for the human brain. This led to some extravagant claims:

> Intuition, insight, and learning are no longer exclusive possessions of humans: any large high-speed computer can be programmed to exhibit them also.
>
> (Simon & Newell, 1958)

> . . . the task of a psychologist trying to understand human cognition is analogous to that of a man trying to discover how a computer has been programmed.
>
> (Neisser, 1967)

The developments listed above created the new discipline of AI. This engineering approach treats organisms as machines, machines controlled by processes. And some of these processes are perceptual. Perception thus offered an obvious challenge to workers in the new discipline.

The research to be described did not arise solely from theoretical considerations. The period when AI was forming as a discipline was also a time when important empirical discoveries were occurring in the study of perception and related areas. We shall summarize some of the most important of these to show what sort of knowledge was available to Marr when he started to build his theory of vision. Four examples will convey the quality of this empirical work.

Receptive fields in the visual cortex

In Chapter 5, reference was made to work by Hubel and Wiesel (1962, 1968), who succeeded in recording the electrical responses of living cells in the visual cortex of the cat and the monkey to various patterns of stimulation. One of the most striking and thought-provoking discoveries was that the visual cortex contains cells responding differentially to lines and edges, according to the orientation of these stimuli. This was a remarkable finding, for it suggested that the visual system analyses visual inputs into specific components, and that the mechanisms that do this are 'wired into' the nervous system. It is therefore possible that the perception of certain basic features of the world is unlearned (although subsequent research showed that the activity of the cortical cells can be modified by prolonged experience).

The Julesz random-dot stereograms

Julesz (1960) discovered random-dot stereograms. When these are fused in a stereoscope a powerful illusion of depth is seen. The depth arises because the paired stereograms contain central portions which differ slightly, thus capturing the cue which normally triggers stereopsis: disparity of left and right views. The strange and wonderful thing about the Julesz demonstration is that the disparity is not visible when one scrutinizes the individual stereograms: one appears to be looking at random textures, arrays which contain no hint of form (see Figure 7.1). This proves that the visual system can extract disparity information in the absence of pattern recognition, a remarkable discovery.

Spatial frequency channels in the visual system

Pantle and Sekuler (1968), Campbell and Robson (1968) and other workers studied various visual systems to see how they respond to changes in the spatial frequencies of test gratings. The *spatial frequency* of a grating is simply the number of changes (commonly the number of black and white stripes) it contains per degree of visual angle. It was found that if an observer stares for a time at a particular grating, sensitivity to that grating is temporarily reduced. That this is not a general loss of visual acuity is shown by the fact that sensitivity to other spatial frequencies remains unchanged. A related discovery was that recordings from the visual cortex of the cat reveal the presence of cells which are differentially sensitive to particular spatial

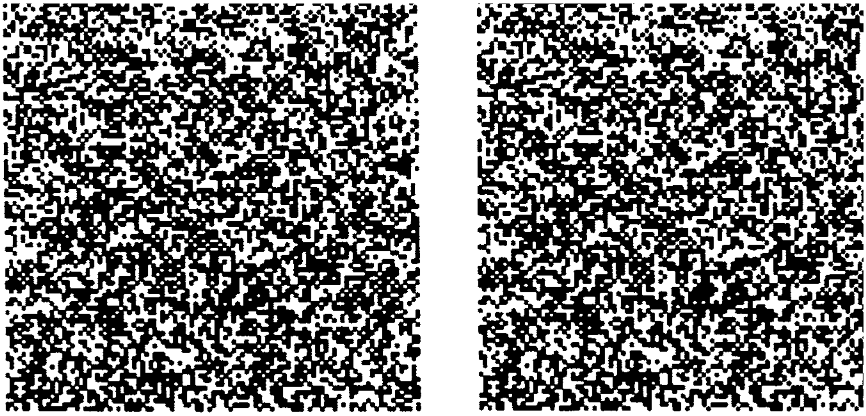


Figure 7.1 Two random-dot stereograms. These are best viewed through a stereoscope. However, if they are viewed either with the eyes crossed or when staring 'through' the page to infinity, the stereograms should fuse. Following fusion depth will be seen to emerge from the apparently random textures. (Thanks are due to Dr D. C. Earle of Exeter University for generating these stereograms.)

frequencies. It began to look as though one could consider the acuity of vertebrate visual systems in terms of tuned channels. Thinking about vision in terms of spatial frequencies led in turn to the use of powerful new techniques of analysis, the most successful of which has been Fourier analysis. These have been very fruitful developments.

Early examples of the computational approach

Land and McCann (1971) and Horn (1974) offered a solution to a classic problem in perception: how does the visual system 'know' that the varied appearance of a coloured surface is a property of the surface rather than its illumination? Their suggestion was that, while the effects of changes in illumination are usually gradual, changes due to a surface's geometry – its edges, boundaries between facets, and so on – are usually abrupt. If a visual system could somehow ignore or filter out the gradual changes, what remained would be information about the characteristics of the surface, rather than its illumination; perception could then be veridical. A solution of the problem was as follows: sample the lightness values of a surface by paired, closely contiguous detectors and record differences in their outputs. Placed on a homogenous surface under an illumination gradient, the differences in outputs will be very small and can be ignored. However, when the two detectors straddle a boundary between two surfaces of different lightness, there will be a large difference in their outputs. In this manner the true surface lightness (or reflectance) properties will be detected and changes due to illumination will not.[2] This was not, of course, an *empirical* contribution, although it was inspired by some striking new colour phenomena which Land had discovered and which caused him to challenge the traditional theories of colour vision.

Describing the genesis of his own ideas, Marr says of the Land and McCann work:

> I do not now believe that this is at all a correct analysis of colour vision or of the retina, but it showed the possible style of a correct analysis . . . gone is any explanation *in terms of neurons* – except as a way of implementing a method. And present is a clear understanding of what is to be computed, how it is to be done, the physical assumptions on which the method is based, and some kind of analysis of the algorithms that are capable of carrying it out.
>
> (Marr, 1982)

---

[2] The actual algorithm needed to achieve this result is in fact slightly more complicated than this: it requires the cumulative storage of ratios of successive output differences, together with rules for deciding which surface in an array can be designated as a standard white.

As we shall show, the *style* of explanations in perceptual research was something that Marr attempted to change. Marr's contribution to the study of visual perception will now be described.

Marr's theory of vision and his programme for research

One of Marr's acknowledged contributions lies in his attempt to clarify our thinking about information-processing systems. In what follows, the reader should keep the following points in mind: (1) the term 'information' is used here more loosely than in the technical sense (where it is related to reduction of uncertainty); and (2) 'information-processing' is not mere transduction of energy. For example, a telescope changes light by magnification, but magnification is not an informational process – the image is simply a linear transformation of the object. As nothing else is really changed by the magnification, we cannot apply informational measures to this effect. Note, however, that we can see things through a telescope that are invisible to the naked eye. This is because the eye itself is a non-linear device. So when we consider the eye and the telescope *as a single system*, it becomes appropriate to use informational concepts and measures.

Representations and descriptions

Marr's definitions of these terms are as follows:

> A representation is a formal system for making explicit certain entities or types of information, together with a specification of how the system does this. And I shall call the result of using a representation to describe a given entity a description of the entity in that representation.
>
> (Marr, 1982)

These definitions, which appear quite abstruse at first glance, may be clarified by a simple example. From a satellite photograph of a country we draw an outline map. Suppose that the satellite's optics have great resolving power and that they can assess the maximum height in each 10 square kilometre area of the terrain, allocating a numerical height value to each. We can now select and add to the map all points with heights of 200 metres, 400 metres, and so on, each represented by a coloured dot. Joining dots of a particular colour (representing a particular height) with straight lines will generate a crude contour map of the country. Next, data are obtained concerning the distribution of people in the country and a single dot is printed to represent each thousand of them. Lines can be added to represent roads and rivers. Finally, we add some markings to represent, say, birth rates in various regions.

A map of the country has now been created from which certain interesting conclusions might be drawn; for example, that more people live in valleys, roads wind round hilly areas, and people living on high ground are more fecund. Now although we have described a single map, it is obvious that five distinct *representations* (outline, height, population density, rivers, roads) have been used to arrive at five *descriptions*: the map can be thought of as formed from five superimposed transparencies.

That was an imperfect explication of Marr's definitions, for the procedures we imagined were not ones which can be done by machines; they would normally be carried out by geographers. However, the important point stands: we have used *symbols* to represent things or events.

Three levels of understanding information systems

Marr's own example of an information-processing task is one that is definitely performed by a machine. (From now on all references are to Marr, 1982, unless stated otherwise.) Marr describes a cash register. It is at this point that he introduces the distinction between the three levels of explanation that, he insists, must be kept separate in our thinking about any informational process. This distinction pervades the whole of *Vision* and may well be one of Marr's enduring contributions. The three levels that Marr distinguishes are: (1) the computational theory; (2) the algorithm; and (3) the hardware implementation.

The computational theory

Marr asserts that at this level of enquiry we must ask, 'What is the goal of the computation, why is it appropriate, and what is the strategy by which it can be carried out?' In the example of the cash register, the function of the machine is clearly to add sums of money. And it is this procedure of addition that brings the machine within the class that we define as information-processing devices: several subtotals may be 'compressed' into a final sum from which they cannot be recovered, so the process is not a linear translation or transduction. In this example, the computational theory is simply the rules of arithmetic. That is to say, it should not matter in what order we enter data into the cash register; if we enter a zero sum, then the total should be unaffected; and so on. We are describing what the machine achieves and also the constraints upon it. These constraints allow the processes within the machine to be defined. At this stage we may be completely ignorant as to *how* the machine does its arithmetic.

The algorithm

'How can the computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?'

The input to the machine is known (key entries which represent sums of money in decimal notation), as is the output (total sums of money displayed in decimal notation), but what has the machine actually done? The machine could have translated keyed entries into electronic or mechanical equivalents of decimal quantities. Or entries might have been translated into binary form (this would be very likely if the cash register was linked to a large computer network). So there is something to discover about the *representation* of the data that the machine will process.

Knowledge of the representation may prompt guesses about the algorithm or formula used within the machine. Clearly, the algorithm chosen will depend in part on the nature of the representation. For example, if the machine is operating according to binary arithmetic, then the algorithm must include some procedure that will change entries from decimal form prior to adding, and then re-convert before displaying totals.

The hardware implementation

'How can the representation and algorithm be realized physically?'

A machine such as a cash register is not fully understood until the implementation is known. We know the computational theory of the machine (the rules of arithmetic) and can form hypotheses as to how input and output are represented and how the correct answer is attained. But there is still an area of ignorance: how does the machine actually work? It might contain interlocking cogs, like an old-fashioned mechanical calculator. Or it might assign voltages to particular numbers using thermionic valves. Or (and most probably) the machine might function by the operation of a series of electronic switches – devices that can represent one of two possible states. The question concerns the hardware of the machine, the nature of its component parts and how they operate.

Many readers, including those with some previous knowledge of perceptual theories, may find these ideas rather strange. As they are central to Marr's approach, it might be useful to pause at this stage and offer a short summary and another illustrative example before proceeding to describe more of his work.

The starting point for seeing is the image on the retina; the end point is our awareness of the world. We seem to have a picture of the world available to us whenever our eyes are open, and we call this 'seeing'. But the truth is that light stops at the retina. There can be no actual picture in our heads, only neural activity. It follows that this neural activity is representing the world *symbolically*, and we must therefore strive to understand this symbolic process or processes. Marr argues that symbolic representations of various aspects of the world, initially obtained from the retinal image, are combined into the descriptions that we call seeing.

Marr suggests that the most rigorous way in which to conduct research is to ask a series of systematic questions arising from the computational approach. Let us consider as an example the perception of contours. In this case the appropriate sequence of questions would be as follows:

- Why is it important to be able to perceive contours? If the visual system can extract them from the visual image, what use is this to the perceiver? In other words, what of importance in the real world correlates with contours in the visual image? Why should the visual system work to make them explicit? How might contours be represented symbolically in our heads? If we can see a contour, is it likely that this has arisen from an edge – a feature that reveals discontinuities between the surfaces of different objects? It is clear that this last question matches Brunswik's concern over the ecological validity of cues. It is equally clear that to answer the questions requires knowledge about (a) the visual system, (b) the purposive aspects of the perceiver's behaviour, and (c) the nature of the real world, which sets constraints upon our theory. This last point is very important. We need to know, for example, how many types of edges there are in the world. When edges form junctions, in how many ways can this be done? What is the relationship between the inside and outside angles of, say, a transparent object? Thus, in this case it is necessary to think about topology.
- When the preceding questions have been answered, it is possible to think about the algorithm. In the present example, we start with the retinal image, which is a set of light intensities spread across part of the retina; this is the input. The output must be a symbolic representation of lines or edges appearing in conscious experience. Now the question arises as to how a process operating on the retinal image could deliver contour information.

The example we have chosen is a relatively easy one, for quite a lot is known about contour perception. A successful algorithm would utilize a well-known property of retinal cells, namely the ability of some cells to inhibit the action of others. Developing this idea suggests that it would be useful if contour perception involved processes that did not pass on information to later stages in the visual pathway from areas in which retinal illumination was homogenous, or even graded in intensity, but which produced outputs in response to rates of change within gradients (the second differential of intensity). This would 'extract' the relevant contour information. We would now start to think about possible excitatory and inhibitory fields in the retina to see whether they could respond to entire edges, and it would be necessary to suggest plausible rules by which these fields could interact.

It is possible to test hypotheses concerning algorithms of the sort described above using electronic circuits, designed so that the components simulate mutual excitation and inhibition. This can reveal whether the circuits can respond in the hoped-for manner to, say, the second differential of a brightness gradient. This can be done in two ways, (a) by actually building assemblies of photo-detectors and electronic components, or (b) by computer simulation. Many ideas in artificial intelligence have been tested in this way.

- Having designed a plausible algorithm, it is now necessary to consider how it could be put into practice. In the present case, independent evidence strongly suggests that it is the retinal ganglion cells that initiate the process of contour extraction. In fact, inhibitory relationships have actually been demonstrated among these cells. It would be of obvious interest to observe the activities of large groups of such cells, but technical problems make this impossible at present.

This, then, is the way in which Marr believes we should approach the task of understanding vision. Lack of progress in the past has often stemmed from failure to ask the right questions about perceptual systems, or from confusion between the three levels to which research attention can be directed. As a test of the reader's understanding of Marr's point, we offer the following challenge (first put to me by my colleague Dave Earle): apply the computational approach to an ordinary lock. What is the computational theory of locking? What would be an appropriate algorithm for the lock? In what way(s) might the algorithm be implemented? Those who find it easy to answer these questions have certainly grasped Marr's argument.

The stages of visual perception

Early in his book, Marr describes how his thoughts on vision developed until he reached his most important insight. What was required, he realized, was '. . . a theory in which the main job of vision was to derive a representation of shape'. Vision can do much more than this, but informing the perceiver about brightness, colour, texture and so on, is secondary to deriving a representation of shape. The problem, then, is to discover how vision derives reliable information concerning the shapes of objects in the real world from information contained in the retinal image.

Marr's theory is that perception proceeds as an information-processing system and that this system is organized into successive *stages*: it is unlikely that reliable or stable conclusions about objective shape could be derived in a single step. Marr also uses his knowledge of computer science to formulate a guiding principle, *modular design*. This principle simply states that when developing computational systems, it is wise to break down the computation into component parts, which should proceed as independently as possible. The reason for this is that if part of a system goes wrong and this part interacts strongly with others, debugging the complete system becomes a formidable problem. Marr's hunch is that many of the processes of vision are modular, and for this important reason.

We shall now outline Marr's analysis of the stages of visual perception:

- *The image*. The 'function' of the retinal image can be defined as representing intensity. The image is a spatial distribution of intensity values across the retina and is the starting point in the process of seeing.

- *The primal sketch.* The function of this stage of vision is to take the raw intensity values of the visual image and make explicit certain forms of information contained therein. The most important information concerns the spatial or geometrical distribution of intensity changes and the manner in which they are organized. The types of information that are becoming explicit at this stage are such as to afford the possibility of detecting surfaces.
- *The 2½D sketch.* At this stage of the visual process, the orientation and rough depth of visible surfaces are made explicit: it is as if a 'picture' of the world is beginning to emerge. Note, however, that at this level what is emerging is organized with reference only to the viewer; it is not yet linked to a stable, external environment.
- *The 3D model representation.* Here shapes and their orientation become explicit as tokens of three-dimensional objects organized in an object-centred framework, that is to say in a manner that is independent of particular positions and orientations on the retina. By this final stage of vision the perceiver has attained a model of the real external world.

Applying the theory

The reader who can see the potential rigour and clarity afforded by Marr's ideas may yet wonder how the approach actually works, how Marr moved from verbal description to a scientific attack on the problem of vision. Selections of Marr's work will now be examined in more detail to convey the style of the computational approach. The first topic is the primal sketch.

Work on the primal sketch

The starting point for this early stage of vision is the array of intensities represented in the retinal image. Marr's theory holds that certain *primitives* or *place tokens* are derived from the image. These are: zero-crossings (which will be explained below), edges, bars (which can be thought of as pairs of parallel edges), blobs (the ends of bars or small clusters of dots), terminations (of edges or bars), edge segments, virtual lines, groups, curvilinear organization, and boundaries.

The development of the primal sketch begins with the derivation from the spatial retinal array of primitives that can be thought of as *tokens*. The idea of tokens is very important in this approach. To explain this a little more fully, consider a technique frequently employed in television commercials, the reverse zoom. A typical advert starts with a shot of a group of individuals. Because we see these as individuals, each must have been assigned a visual token. Now the camera zooms out and we start to see that the people form various groupings. Finally, we see, from a great height, that the people are arranged as letters in a word (the advertised product, etc.). Seeing each letter as a coherent whole implies that it too must be represented in the visual system, and thus, in Marr's terms, another token has been created. Of course, the tokens formed in the visual system are 'really' neural events. When Marr or other theorists in AI use actual visual tokens to illustrate the successive processing of images, this is argument by analogy.

During the development of the primal sketch, groups of adjacent tokens having a common orientation are replaced by 'level one' tokens representing this orientation. Then, if there are whole groups of similarly orientated level one tokens, these are used to construct boundaries between parts of the full primal sketch. There is nothing mysterious about these notions. Remembering our illustrative example of a map, it is obvious that when a sufficient number of concentric contours occupy a given region, they could all be replaced by a single purple patch to indicate a mountainous area – cartographers do this routinely – the patch would form a token, like those formed for the letters and words in our television commercial example.

A question arises as to how the initial primitives of the primal sketch are actually extracted. The attempt to answer this question leads Marr to a most impressive piece of work and demonstrates the advantages of his background in AI. We shall concentrate upon the primitive known as the *zero-crossing*.

When a photograph or an actual image (such as that on a television screen) is scrutinized, it is obvious that important information about the shape and orientation of objects comes from edges, contours and boundaries: that is, from areas in the image where intensity values are changing rapidly. Clearly, then, it should be important to represent these portions of the image in the primal sketch. But what sort of process can take intensity values as inputs and deliver tokens representing lines, edges and so on as outputs?

The first step in the required processing is to smooth the image. Light is 'noisy' and it is important to minimize this noise before doing further processing. A process that will do this is *convolution*. One way of convolving an image is to choose a particular location (often defined as a *pixel*) and then apply weighting functions so that, following convolution, the intensity value at the location is replaced by the weighted sum of itself and adjacent pixels. The process is then repeated at all positions on the image. In starting to think about this problem, it occurred to researchers that one method of weighting would be to apply a Gaussian distribution to successive portions of the image, such that regions under the centre of the Gaussian were weighted strongly, those in the periphery less so. The width of the Gaussian can be controlled by adjusting its standard deviation: the wider the distribution, the greater the degree of smoothing of the image and the smaller the range of spatial frequencies transmitted. A narrow Gaussian will pass more high spatial frequencies. (As we shall see later, combining a positive Gaussian filter with a negative one – in which the weights are subtractive rather than additive – yields valuable results in this method of processing.) The type of technique we have described is in fact widely used in applied situations such as computer image enhancement.

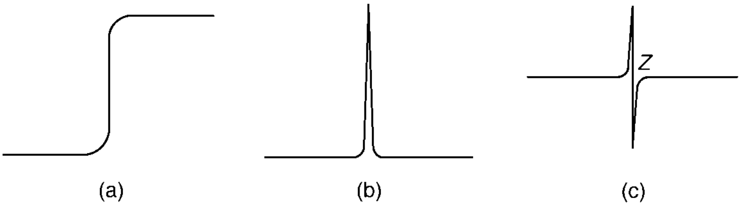


Figure 7.2 The extraction of a zero-crossing (see text for an explanation). (From *Vision* by David Marr, copyright © 1982 by David Marr. Reprinted by permission of Henry Holt and Company, LLC.)

Now examine the graphs drawn in Figure 7.2 (adapted from Marr, 1982). In this figure, (a) represents a change in intensity in part of an image – this is the input to our hypothetical process. (b) Is a representation of *the first derivative* of the intensity change. By this is meant simply that the intensity curve in (a) has been re-plotted in (b) to show the rate of change of intensity: this is clearly zero at the start, then rises rapidly before falling to zero again. (c) Is *the second derivative* of (a) and represents the rate of change of (b). This amounts to saying that if we examine curve (b) we can find points when it too is changing rapidly; plotting these changes in (c) reveals that the curve is initially at zero, then rises positively (upwards) quite rapidly to a peak before diminishing (in a negative direction) with equal rapidity. In moving from its positive to its negative peak, the graph crosses the horizontal zero axis. This, in Marr's theory, is the primitive we require: the *zero-crossing*. We have moved from an image (or part of one) to a representation. (In a later part of his theory, Marr guesses that the *sign* of the zero-crossing may be represented; this would give additional power to the extraction process.)

The computational theory has suggested the zero-crossing as a primitive. The next stage is to construct an algorithm – a set of rules by means of which zero-crossings may be extracted from images. Marr's suggestion takes the form of an *operator* or mathematical function; $\nabla^2 G$, where ∇^2 is the Laplacian operator

$$\nabla^2 = \frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2}$$

and G is the two-dimensional Gaussian distribution:

$$G(x, y) = \exp\left(-\frac{x^2 + y^2}{2\pi\sigma^2}\right)$$

The reader for whom this sort of formula is very unfamiliar should not feel discouraged: Marr is simply describing a mathematical process that will convert intensity changes in a two-dimensional image into zero-crossings, rather as in the one-dimensional case illustrated in Figure 7.2. What happens is this: ∇^2 the Laplacian operator, extracts the second differential information that we require. G, the two-dimensional Gaussian distribution, will blur the image by controlled amounts. The result of using these two formulae simultaneously is that zero-crossings can be extracted over a range of spatial frequencies. Put another way, ∇^2 is a band-pass filter.[3] As we shall show, this is very important.

It is mathematically certain that the above operations will do the required extraction, but, of course, these are ideal, abstract formulae. There is, however, a filter which closely approximates the ∇^2 operator: this is known as a difference of two Gaussians (DOG). The performance of DOG filters is known and they have certain advantages for the present purpose, one of which is that, like the ∇^2 operator, they too can be tuned to different scales. The importance of tuning can be demonstrated very easily. The reader should stare at this page with partially closed eyes. Then the page will be seen as an area of brightness that differs from the surface on which it is lying. Similarly, the paragraphs on the page can be seen as blocks of dark grey against the white page. But on opening the eyes again it can be seen that there are discontinuities in brightness operating at a much finer scale: one is aware of lines of black print and spaces between words and letters. This shows why a successful filtering process attempting to capture zero-crossings should be capable of operating at more than one scale. More importantly, by using spatial frequency filters tuned to different scales and comparing their outputs, the chance of detecting an actual edge – one that is present in the external world – is greatly increased. In fact, Marr and Hildreth (1980) have proved that if several different spatial filters over a contiguous range of sizes indicate the existence of a zero-crossing in the same position in the image, *then this must arise from a single physical cause, for example, an edge in the world.* This is a most important finding. Knowledge of the physics of the real world is confirming major theoretical assumptions.

It is, however, possible that the reader is experiencing some puzzlement at this point. Why, when ∇^2 will do the required extraction of zero-crossings, should emphasis suddenly switch to DOG filters? The answer is that when considering a suitable algorithm in this attempt to solve a problem, Marr is remaining aware of the next stage he must deal with: the implementation. Now the formulae summarized by the symbols ∇^2 are, as we have shown, highly complicated. It is unlikely that there are neural mechanisms in the early stages of the visual system which can perform directly the advanced mathematics required. However, from what is known about the ganglion cells of the retina and their receptive fields, it is entirely reasonable to

---

[3] Low-pass and high-pass filters transmit low and high spatial frequencies, respectively. A band-pass filter transmits a particular range of frequencies. Used alone, G would act as a low-pass spatial frequency filter.

suppose that these receptive fields could yield outputs, one type of which resembles positive Gaussian weighting functions, another negative ones. Thus, the DOG has the advantage over ∇^2 in that one can begin to see how its job could be done by neurons.

An important advantage of the AI/engineering approach to vision is that it is possible at this stage to make a powerful indirect test of the theory so far. It is possible to substitute photographs of real scenes for retinal images and process them through actual filters. Then, by examining the outputs of these filters, one can see to what extent lines, edges and so on, have been made explicit and whether the shapes in the photographs have been separated. It is also a simple matter to assess filters other than DOGs to compare their performances (if the reader will look back to the computer-processed photographs in Figure 4.12, it will be seen that a DOG filter can be approximated by superimposing a mask containing positive and negative weighted regions over a pixel array).

Marr's publications contain numerous illustrations of such filtering procedures. We must point out, however, that it is not yet *proved* that zero-crossings are computed by the visual system, but this is not Marr's main concern at this stage. After finding an appropriate computational theory and related algorithms, it becomes necessary to consider the hardware implementation – to look for neural devices that will extract zero-crossings and other primitives. From what is known about the neurophysiology of the retina and visual pathways, it is obvious (as was stated above) that this search should concentrate upon systems or cells having inhibitory capabilities, such as retinal ganglion cells and others in the lateral geniculate nucleus that show receptive field properties.

Receptive fields can be organized in various ways and have various shapes. Marr's guess concerning the cells delivering information about zero-crossings is that they are the retinal ganglion and lateral geniculate cells known as X-cells: in particular, those having On-Centre/Off-Surround organization (firing when the centre of the receptive field is stimulated, inhibited when the surrounding portion of the field is stimulated), and the Off-Centre/On-Surround cells (having the opposite type of organization). In a striking demonstration of the probable truth of this part of the theory, Marr displays, simultaneously, the outputs of DOG filters to lines and edges and the outputs of actual X-cells responding to the same stimuli. The similarity between these outputs is remarkably close.

An application to stereopsis

Following that description of work on the early stages of vision, we shall show more of the rigour and power of the computational approach by describing an attack on a second problem. We shall now give an outline of a possible solution to the problem of stereopsis.

The reader who wishes to experience at first hand the high quality of

Marr's work should consult the relevant chapters of *Vision*, or the original research paper by Marr and his collaborator, T. Poggio (Marr & Poggio, 1979). Beware, though; this is very difficult material and it can take several readings before one feels confident that the arguments have been mastered. It is well worth the effort, however.

Stereopsis is a term having two meanings. First, it refers to that extra sense of solidity and depth that is experienced when using two eyes rather than one – an experience that is confirmed by the superiority of binocular depth discrimination. It is a cue to the relative rather than absolute distances of objects. Second, stereopsis is triggered when two slightly different views of a scene are viewed in a stereoscope. In this case, the two flat patterns are inducing an illusion of depth which can be manipulated during attempts to gain an understanding of normal stereopsis.

Classical work showed that the basis of stereopsis must lie in the difference between the left- and right-eye images. When these fall onto corresponding areas of the two retinae, disparate images will induce an illusion of depth. That this *disparity* is a necessary and sufficient cue for stereopsis is demonstrated by the ability of random-dot textures to induce depth when the actual disparity is hidden, as it is in Figure 7.1. This figure shows that the problem of stereopsis is a formidable one: how is depth assigned to such stimuli? And how is this possible in the absence of all the familiar monocular cues to depth, such as size, perspective, and shading, and also the absence of recognizable form (e.g., one cannot match a detail such as a branch of a tree in the image of one eye with the same detail in the other eye's image)?

The value of the computational approach becomes evident at the start. Simply to state, as many have, that disparity is the basis of stereopsis, is insufficient as an explanation. Marr goes deeper, and begins by asking two questions: how is disparity *measured* by the visual system, and how is it *used*?

We shall outline Marr's attempt to deal with the first of these questions, the measurement of disparity. Marr's analysis of the situation at the two eyes during binocular viewing leads him to recognize two major related problems (he was by no means the first person to describe them). Although the problems are easy to describe, it has taken over 100 years to find plausible solutions.

Figure 7.3 is a diagram of the situation when a person looks at a row of, say, lights at a fixed distance from the two eyes. The eyes have been drawn symbolically in order to display the patterns of stimulation at the two retinae. As the process of vision starts at the retinae, stereopsis must take the inputs there as the vital information concerning the locations of the lights. Thus, the solution as to where the lights are located must be found in the relationships (or matches) between the two retinal patterns. But examine the situation closely: by drawing rays from each light to the two eyes we create the crossover pattern shown in the Figure 7.3. This figure is important because it reveals that *the patterns falling upon the left and right retinae are*

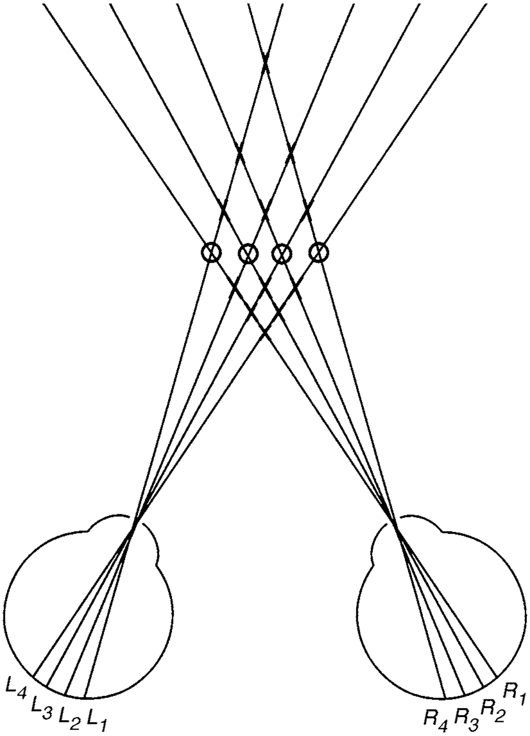


Figure 7.3 The false location problem in stereopsis. The circles represent four coplanar objects viewed from the two eyes. L_1–L_4 and R_1–R_4 are corresponding positions on the left and right retinae. Crossings represent false locations: positions from which patterns of stimulation at the eyes could arise that are identical with those arising from the four coplanar objects. The problem is to account for the fact that the visual system can solve this problem and avoid false locations.

not uniquely determined by the configuration of the lights, but could have been caused by any of a number of alternative configurations: those represented at the various crossover positions in the diagram. We have used a small number of stimuli in this illustration; the number of false locations, as they are known, grows exponentially with the number of stimuli and would be huge in many real-life situations.

The two problems then are: (1) how are parts of one image correctly matched with parts of the other; while (2) avoiding those matches arising from false locations. In other words, how is the truth obtained from the ambiguous information in the two visual inputs? Marr realized that if he could account for the stereoscopic depth induced by random-dot displays, which are meaningless and contain no other depth information, then an account of 'normal' stereopsis would follow quite easily. The dot displays contain, in a sense, 'pure' disparity. A world sprayed with dots is also simpler to consider.

As would be expected, the computational approach to stereopsis involves three levels of discourse, the first of which is the computational theory.

The computational theory of stereopsis proceeds as follows. We are considering the perception of a three-dimensional world containing only textured surfaces, and have discovered a fundamental problem: how is a dot seen by one eye correctly matched in the other while avoiding false matches? Marr starts by adopting two constraints set by the nature of the world in which the visual system evolved:

- A given point has a fixed position at any moment in time.
- Matter is cohesive. Surfaces are not arranged in ways that can trick us. They cannot, for instance, suddenly bend or change their curvatures without yielding some clue to such changes.

The formal computational theory begins with three matching rules applied to a textured surface which is in binocular view:

- Black dots can match only black dots. We are considering surfaces that contain only black dots and white spaces, and the rule is simply stating that a point on a surface seen by one eye stays the same when seen by the other.
- A black dot in one image can (truly) match only one dot in the other.
- Disparity, the magnitude of the difference between the left and right eye matches, varies smoothly. Once again it is being assumed that the world does not (cannot) play tricks.

We now have the beginnings of a computational theory. Can a combination of the matching rules and the constraints suggest a solution to the problem of stereopsis? Marr now presents an interesting analysis of the situation at the eyes when both are looking at the same scene. This analysis is illustrated in Figure 7.4. In this figure the positions of 'descriptive elements' in the left and right images are plotted along the two axes. Horizontal and vertical lines represent lines of sight from the left and right eyes, respectively. Where these lines intersect are possible disparities, that is, positions of matches and false locations as shown in Figure 7.3. The dotted diagonal lines are lines of constant disparity, or positions along which the differences in left- and right-eye views of a surface have the same magnitude.

This deceptively simple diagram is an important part of Marr's proposed solution of the problem of stereopsis. The aspect of the diagram to note is that the distribution of matches and false locations is not chaotic: both are spatially distributed among the two images in an orderly manner. A regularity such as this, existing in the physical world, suggests that here is a source

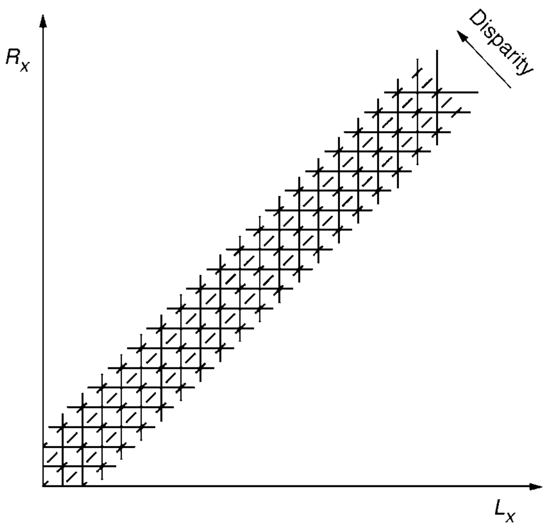


Figure 7.4 The Marr–Poggio analysis of stereopsis. The vertical and horizontal axes plot the distribution of stimuli to the two eyes: where these meet are possible positions giving rise to disparities of view of the same object. The dotted diagonal lines are lines of constant disparity. Rule 3 of the Marr–Poggio algorithm requires correct matches across the eyes to cluster along the diagonal lines. This limits the set of possible solutions and thus a potentially chaotic situation becomes a manageable problem: there are noticeable regularities that could be the basis of the true matching of left and right stimuli during stereopsis (see the text for a fuller account of the stereomatching algorithm). (Reprinted with permission from Marr & Poggio, copyright © 1976 American Association for the Advancement of Science.)

of information. This in turn implies that there should be a solution to the problem of correct matching.

Support for this optimistic conclusion comes when the simple matching rules described above are combined with an equally simple logical analysis of the situation. Remember that by implication there can only be one match on any single line of sight. Now assume that, for example, the density of dots in each image is 20%. In other words, there is a one in five chance that a dot will be present at any location. Now consider the correct plane in Figure 7.4. What can be said about the density of possible matches – of dots which can signify by their disparity that a particular surface is present? Clearly this, too, must equal 20%. But what of the density of possible matches on the *incorrect* planes in the figure? A little reflection shows that as these planes are the wrong ones, dots will match only by chance. This gives us the same probability of matching that would occur if two transparent sheets, each with a (random) dot density of 20%, were superimposed. This probability is calculable and is simply the product of the two individual densities (more simply, their probabilities, which are the inverse of the densities): p = 0.2 squared equals 0.04. Hence, provided that the difference between the density (or probability) value in the 'true' situation is detectably different from those in all remaining situations, the matching rules will yield a unique solution. The question now is whether the visual system (or a mathematical algorithm) can profit from the orderliness that has been demonstrated. With this in mind, Marr continues by asking what could be the input to the stereo-matching process required by the computational theory. He suggests that it must take the form of zero-crossings from filtered images. Evidence outlined earlier suggests that the visual system has various tuned spatial frequency channels. If the first attempt to find stereo-matches uses inputs from the larger (lower spatial frequency) channels, this will have two benefits. First, the number of possible matches (and false matches) is reduced because the search through the arrays is coarser. Second, finding some evidence of corresponding matches in an array can direct eye movements so that finer and finer spatial channels can be used for further searches.

The algorithm

Marr now considers various ways of designing an objective procedure that will lead to a solution of the problem. His proposal is that the ∇^2 operator is applied to each image and the zero-crossings (mentioned earlier in this chapter) are extracted. The operator will act as a bandpass filter.

As the algorithm is developed, a surprising result emerges. It is of the nature of bandpass filters that the zero-crossings which form their outputs cannot occur at less than particular spatial separations. Thus, the probability of finding a match between zero-crossings in each eye can be calculated, as can the probability of finding false correspondences. Marr shows, in a plausibility argument, that in a variety of situations, the desired matches will far outnumber false matches when a particular disparity value close to the truth is being evaluated. If a new disparity value is assessed (and, let us say, is false), then the ratio of correct to incorrect matches will fall dramatically. All that one needs to add to the algorithm is the ability to know a good situation, in terms of successful matches, from a poorer one.

Although we have managed to avoid the let-out phrase, 'It can be shown . . .', the reader may feel a sense of unease at this point. Does this technical claim that something will work dodge a real explanation as to *how* it might work? Two additional points may help convince the sceptical reader.

First, it is important to remember the constraints that the computational theory made explicit. Think about the two eyes looking at a flat surface. The two views must be slightly different. The size of this difference (the disparity between the two images) contains the information as to where the surface is in depth, and this is what the visual system is trying to calculate. But remember that in normal viewing there really is a surface there before us. An element of the surface seen by one eye is actually visible to the other – it cannot suddenly move or disappear. And if the selected disparity value is correct, it will be correct over a major portion of the display, for a physical surface does not move away or change in an instant (this is one of the constraints adopted above). If the visual system makes a wrong calculation as to the disparity value for part of the surface, then not only are the matches at that point wrong, they will be generally wrong all over the image. In other words, it is because of plausible assumptions that can be made concerning the real world that certain procedures can be guaranteed to have a high rate of success.

There is a second reason for thinking that Marr's algorithm might be correct: it works. A connectionist network designed according to the algorithm that we have outlined can solve the Julesz stereogram displayed in Figure 7.1. We do not need to take this part of Marr's work on trust; we can see it work in practice. The algorithm is sufficiently explicit and powerful to allow the computer to find the hidden disparity in random-dot displays and represent the depth difference associated with it. Put very crudely, the network functions by comparing portions of each of the stereograms. Only if a particular unit receives inputs from two identical features (two white dots or two black ones) will it become active. If then units representing the same lines of sight from the two 'eyes' receive inputs from parts of the display that would represent different disparities, inhibition passes between them to check this incorrect state of affairs (which would violate the uniqueness constraint described above). After a number of iterations, the network settles upon the correct answer: the part of the displays seen in depth by human observers has been delineated. This is an impressive and convincing demonstration.

The neural implementation

The final stage of what Marr considers to be a satisfactory explanation (or model) of stereopsis is the implementation. What sort of neural hardware could carry out the operations contained in the algorithm? Marr admits that the search for a plausible neural implementation of his theory may be premature, given our rather hazy knowledge of the neurology of some parts of the visual system. He does, however, offer some hypotheses as to how his model could be implemented.

The ways in which cells in the visual system could mimic the operations of the ∇^2 filter have been dealt with earlier. The detection of zero-crossings can be achieved by simple logical gates, which neural cells can mimic by suitable interactions between excitatory and inhibitory processes. In stereopsis it is necessary to combine binocular information about zero-crossings and their signs (positive-going/negative-going) in order to match, say, black dots to black dots. Once again, logical devices (in this case AND gates, which fire when both possible inputs are active) are capable of doing the required work, with each gate having as one of its inputs the difference between the left and right eye inputs at the position of whichever zero-crossing (left eye or right eye) is chosen as the starting point for the comparison; the zero-crossing forms the other input to the gate. Exactly where in the visual system these AND gates will be found is uncertain. Marr guesses, on the basis of published micro-electrode studies, that the proposed disparity detectors may lie in Area 18 of the visual cortex. The fine resolution of depth that is possible at the limit of stereo-acuity may be based, at least in part, on the activities of granular cells in layer ivc Beta in Area 17 of the visual cortex. These would use inputs from the high resolution spatial outputs from the filtered images.

Further aspects of Marr's work

The understanding gained from the attack on the problem of stereopsis allows Marr to extend the computational approach to other areas of perception. Further chapters in *Vision* contain interesting discussions of directional sensitivity, the perception of motion, shape, and contour, lightness and brightness, shape from shading, and, finally, the perception of three-dimensional objects when perception moves from viewer-centred to object-centred frames of reference. In terms of the model of vision, the processes have finally arrived at a description of the objective world.

Many of the later parts of *Vision* are very interesting. In a discussion of the recovery of shape from silhouettes, for example, Marr's careful analysis of the stimulus situation allows him to predict when a silhouette is a reliable guide to shape and when not: it is reliable when the portions of a surface generating the silhouette are in the same plane, not otherwise. Reading these sections, one feels that few can ever have thought so analytically and deeply about the nature of the three-dimensional world and the ways in which it gives rise to visual images.

Towards the end of his book, Marr attempts to outline how we perceive three-dimensional shapes. An important part of this work concerns the ways in which the visual system uses *canonical forms* in a *modular* organization (i.e., split into different parts). As an example, consider the attainment of the three-dimensional representation of a human being. One possible canonical form that would be useful here is the cylinder. Following the principle of modular organization, an initial cylinder could be constructed to represent a person, provided the visual system could first decide upon the direction of the principal axis: in this case, from head to feet. This would be a self-contained unit in the shape description. It is possible to enrich the description, using the same canonical form to represent the head, the torso, the arms and legs by smaller cylinders. Next, the arm could be represented by a set of cylinders, one of which is the hand. Finally, the hand cylinder could be elaborated into a set of cylinders representing the wrist and the fingers. The same canonical form, the cylinder, has been used throughout, but successive applications at different scales yield descriptions that are increasingly 'lifelike'. In terms of the computational approach, a three-dimensional model of a human being has been attained.

It is clearly inappropriate to attempt a detailed description of the whole of Marr's work. We shall, however, state an opinion, which is that, despite the ingenuity of his reasoning, Marr's ideas on the perception of objects are less impressive and convincing than the earlier chapters of *Vision*. This is hardly surprising: the problems are much more formidable. Thus, for example, the account just given of Marr's approach to the perception of three-dimensional shape gives one the impression (and this is not true in other parts of Marr's work) that, while his ideas seem quite plausible in terms of machine recognition of objects, the evidence that this is how a living visual system might function is less than compelling.

An appraisal of the computational approach to vision

It would be unwise to offer too confident an evaluation of the computational approach to perception. The work of Marr and his colleagues was carried out in laboratories specializing in AI research. It has taken some time for the ideas of these workers to become widespread in psychology and physiology. As we have been able to demonstrate, many of the concepts and mechanisms invoked to explain perceptual phenomena are complicated. The language is not one that is familiar to many who work in other disciplines. The mathematics is occasionally difficult, and lacking the facilities to use the various operators and filters described by Marr means that some readers have to take his findings on trust. This state of affairs is changing. But it will be some time before every worker in perception will be able to demonstrate computer-filtered images as easily as they can generate, for example, Mach bands, random-dot stereograms, or rotating shadows.

Particular problems

There are, however, two studies which are worth reporting here, as they throw some doubts upon the adequacy of one important part of Marr's computational model. Mayhew and Frisby (1981), who work within the computational framework, present psychophysical evidence from experiments using stereograms. It will be remembered that in Marr's model the raw primal sketch makes intensity changes in the image explicit by using primitives such as bars, blobs, terminators, etc. These in turn are replaced by more abstract tokens, which lead to the achievement of the full primal sketch. In the full primal sketch only two-dimensional projections of objects are represented. The system is not concerned with the extraction of three-dimensional disparity information until the later stage of the 2½D sketch. Mayhew and Frisby used sawtooth patterns as stereograms. In these cases, the depth that resulted is not predictable from knowledge of zero-crossings – the primitive which Marr adopts as the input to a stereo-matching process.

Mayhew and Frisby have published examples in which the positions of zero-crossings are identical in two stereograms (and hence cannot signal disparity) and yet these can induce stereopsis. Thus, in Marr's model one would have to include other sources of information, for example, the peaks obtained from the convolutions, in order for stereopsis to be achieved. More importantly, Mayhew and Frisby present convincing arguments in favour of a model of stereopsis in which disparity is computed much earlier than the 2½D sketch: probably at the level of the raw primal sketch.

Watt (1988) and Watt and Morgan (1985) have also pointed to weaknesses in Marr and Hildreth's work on zero-crossings. They have developed an improved algorithm of greater complexity than Marr and Hildreth's, and one that more closely matches human discrimination. The two references cited above should be consulted for a detailed account of these highly interesting developments.

The author's colleague, Dave Earle, has published evidence that suggests that another part of Marr's thinking may be wrong. Earle used Glass patterns (Glass, 1969; Stevens, 1978), which Marr cites as important evidence in favour of part of his model. A Glass pattern forms when two patterns of dots or other simple shapes are superimposed. The first pattern is typically a random array, the second is some transformation of the first – an expansion or a rotation, for example. Depending upon the transformation, merging the two displays gives rise to an organized pattern having a strong radial or circular appearance. A typical pattern is shown in Figure 7.5.

To account for the organization of Glass patterns, Stevens (1978) proposed that local pairings between adjacent elements are represented by the visual system as virtual lines. Marr adopted this suggestion and made virtual lines one of the primitives of the primal sketch. Thus, Glass patterns must be revealing some of the workings of this stage of vision.

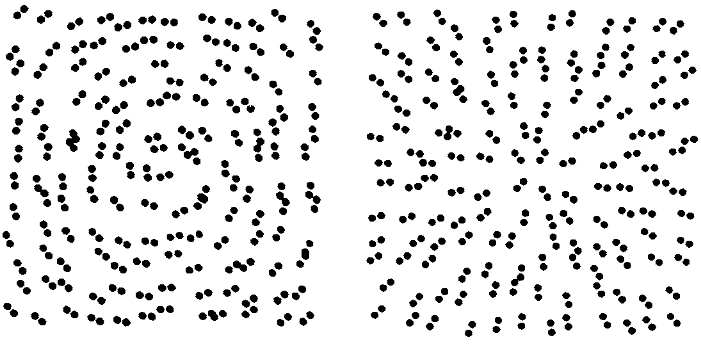


Figure 7.5 Glass patterns. (a) is formed by a superimposition of a random dot array and its rotation; (b) is formed by a superimposition of an array and an expanded version of the same array. (Prepared for the author by his colleague, Dr D. C. Earle.)

Remember that in Marr's model the primal sketch is not concerned with the three-dimensional information in the image. Glass patterns should not occur if the two component patterns are presented stereoscopically, one to each eye. In an elegant series of demonstrations, Earle (1985) has created stereograms in which Glass pattern structure is in fact destroyed by apparent depth; more importantly, he has also designed stereograms in which novel *three-dimensional* Glass patterns arise when no structure is visible in the two-dimensional components. This important finding suggests that if virtual lines are constructed by the visual system at some stage in image processing, the stage must be one in which depth is made explicit. Combining Earle's findings with those of Mayhew and Frisby cited above, we are forced to the conclusion that a detailed but important part of Marr's model has been tested and found wanting.

A new trend in vision research?

The importance of Marr's contribution to our understanding of vision is unchallengeable and his work continues to exert a major influence on visual theory and research. When the search engine, Google, was given the entry, 'David Marr Vision' (in July, 2003), it came up with more than 13,000 hits. This is an impressive figure. There are, however, some trends in the modern literature that suggest that some workers have moved away from any strict adherence to Marr's approach. The best way to detect new trends in any research area is to read all the recent journals, attend conferences, talk with researchers, and so on. However, the present author does not work in the field of artificial intelligence and so must write tentatively as an outsider. The next few paragraphs are intended to show how things in Marr's field of theory and research may be changing.

Earlier in this chapter a description was given of Marr's rule for progress in theorizing. This held that it was necessary to distinguish between the goal of a visual process (the computational theory), the rules by which the goal could be attained (the algorithm) and finally the neural means by which the rules could be realized (the implementation). However, some modern workers, many of whom acknowledge Marr's work in the introductions to their publications, seem to be adopting a new style of working. We shall illustrate this point by outlining some work recently published by Shams and von der Malsburg (2002).

In primates, the primary visual cortex (Area V1) is dominated by the presence of complex cells (see e.g., Mechler & Ringach, 2002). The behaviour of such cells was mentioned in Chapter 4: typically, they respond to edges or bars having particular orientations that fall within their receptive fields. These cells are characterized also by sensitivity to spatial frequency, but are not sensitive to the spatial phase of gratings (Skottun et al., 1991). On the face of it, this is rather puzzling, as it has been demonstrated frequently that phase information (light or dark, black or white, positive or negative) is very important in the perception of complex scenes. Indeed, when the phase information in photographic images is distorted in a random manner, the images become unrecognizable. If phase information is so important in vision, why should the majority of cells in the primary visual cortex be insensitive to this information? Why does this insensitivity not disrupt perceptual processes?

A complex cell's function is to represent something, to provide information for more central regions of the visual system where shape, position in depth, and so on are calculated. So, how ambiguous is the representation offered by a complex cell? It is known from the study of individual complex cells' responses that it is not possible to determine whether the stimulus that led to the response is an edge or a line, or whether it is black on white or white on black. Neither can one tell whether it is a bright bar on the edge of the cell's receptive field, or a fainter stimulus close to the centre of that field. Thus, when considering the output from any single complex cell, there is ambiguity concerning the nature of the input it has received.

Shams and Malsburg modelled complex cells using Gabor wavelet filters. We have already come across the idea of a filter earlier in this chapter when describing Marr's work using difference of two Gaussian (DOG) filters. Basically, a Gabor wavelet filter is a complex sinusoid combined with (or modulated by) a Gaussian function. Neural networks of the type described in Chapter 4 can be used to run a series of such filters. Input images are given to the network to see whether outputs are recognizable versions of the original images (in actual practice, the procedure is very much more complicated than this, but we hope that the general idea is fairly clear).

Gabor analysis can reveal what are known as Gabor magnitudes, and it is known that these are close approximations to the response outputs of complex cortical cells. When Shams and Malsburg ran images through their model they found, surprisingly, that good image recovery was possible; in other words, despite the lack of phase information, the model worked well. What is the explanation of this finding? Shams and Malsburg conclude that when a *population* of simulated complex cells' responses are measured, any ambiguities in the outputs is always less then the ambiguity associated with the output of any single cell. And that when any valid collection of magnitudes is scrutinized, '. . . phase information is automatically and implicitly encoded as well'. And this is the basis for object recognition. Further, the partial loss of phase information could give this part of the visual system a degree of flexibility. That is to say, the small ambiguities in outputs that can be inferred from this research confer upon this part of the visual system a degree of robustness: to alterations in illumination, background, and small distortions of shape in the visual image,

Readers who have persisted thus far may be wondering why this complex technical material has been added to our assessment of Marr's work. The answer is simply this: the type of thinking that led to the above analysis is not like that demonstrated in Marr's own work. It is in fact possible to translate the procedures followed by Shams and Malsburg into Marr's scheme. Marr's computational theory (the goal) in this case would be the extraction of shape information from neural information reaching parts of the primary visual cortex; the algorithm would be the Gabor wavelength function, and the implementation would be the complex cells in the visual cortex. However, this is not the order in which Shams and Malsburg proceeded. They took as their starting point the actual behaviour of complex cells in area V1 of the visual cortex; they then looked at available algorithms and chose the Gabor wavelength function; finally, these researchers showed how their results could deliver the goods, as it were.

There are hundreds of recent studies similar to that outlined above. If we are correct in seeing this as a trend in theory and research, then it is a trend away from Marr's stipulations regarding the best ways of approaching problems in visual perception.

We have gone into considerable detail in the above paragraphs. This was a deliberate tactic. It is hoped that it will illustrate the high degree of precision of thought and experiment demanded of those who wish to test parts of Marr's theory.

We shall now offer some general comments on the computational approach to visual perception as exemplified by Marr's writings.

Workers in artificial intelligence distinguish between bottom-up and top-down processes. Bottom-up processes involve lower-level, more peripheral systems that are relatively autonomous. Their outputs can be fed into higher levels of the system. Top-down processes are well described by their name. To use one of Marr's own examples, it might have been the case that the visual system solves the problem of stereopsis in a top-down manner. A major portion of the image in the left eye (say, the representation of a tree) would be chosen and compared with the right-eye view of the object. Then the branches could be scanned, then the twigs, and finally the leaves, comparisons being made at all these scales. This process would profit from, and be guided by, the perceiver's knowledge of trees. It would be a top-down process. (In this case, however, our ability to see depth in meaningless random-dot stereograms tells us that stereopsis must in fact make considerable use of bottom-up processes.)

When this distinction is applied to Marr's model it is clear that he was guided by certain top-down considerations. The basic idea that the *goal* sought by a perceptual process should feature importantly in any explanation of that process obviously involves a degree of top-down modelling. However, much of Marr's detailed work concentrates upon what are clearly bottom-up processes, such as, for example, those involved in contour extraction and texture discrimination.

Marr is generally most convincing when he speculates about bottom-up processes. We are told what is required to achieve some desired result, what the inputs and outputs to the necessary computations must be, and which neural mechanisms could do the necessary work. It is always satisfying to work through expositions of this kind. They contribute importantly to the high quality of *Vision*.

It is when Marr speculates about processes involving knowledge of the world and acting in a top-down manner that his work becomes less convincing. For example, do we carry in our heads those canonical shapes, such as generalized cones and cylinders, which he postulates as the primitives for three-dimensional shape perception? How could we test this possibility?

The computational approach arose from within artificial intelligence research. Fundamental criticisms of AI have been made by those who doubt whether it is in principle possible for machines to simulate human processes such as thinking and perceiving. Some doubts concerning the machine as an analogue of the mind have been reviewed in Chapter 5. There are a few others to which the reader's attention should be drawn.

As an example, consider some of the arguments advanced by Gregory (1995). Gregory makes the point that, although digital computers work by carrying out computations, analogue computers do not. If in fact the brain works as an analogue device, then it does not compute and is therefore not a computer. Thus, the fact that Marr can *describe* certain visual processes in terms of algorithms does not prove that they work by algorithms. We know that perception is rarely entirely accurate but that it is fast, even though its hardware is relatively slow. Facts such as these support Gregory's belief that analogue models will more accurately capture the ways in which the visual system operates. If he is correct, then Marr's assumptions concerning the ways in which the visual system computes solutions to problems may be ill-founded.

A related point is worth making here. Simulation is an important tool in science and has been used for centuries: think for a moment about the ancient models – lovely brass constructions – of the planets circling the sun. These work by means of cogs and shafts, while the planets move under the influence of gravity. In the School of Engineering in the present author's university is a large model of the Exe estuary. Every so often, the model is filled with water in a manner that simulates the incoming tide; after a time, the model drains and one sees the behaviour of the outgoing tide. It is a very attractive display. However the Exe, like other tidal estuaries, behaves as it does because of the gravitational pull of the sun and the moon; the simulation, valuable as it is, does not capture the ultimate truth concerning the behaviour of the tides. There are some who believe that some of Marr's simulations, impressive as they are, may not capture the basic truths of how visual perception actually works.

This chapter on Marr's work must end on a note of appreciation. There seems little doubt that, whatever the eventual fate of the computational approach generally, or of Marr's contributions in particular, readers of *Vision* have been in contact with an interesting and original mind. The quality of Marr's work has exerted a major influence upon theorizing in perception. The idea of different levels of explanation of processes is a powerful one. Realizing that a newly discovered neural mechanism can never provide a sufficient explanation of any aspect of perceiving until one has asked the computational question – what *function* does the mechanism subserve? – is a definite gain in theoretical sophistication. So, too, is the idea that the real world exerts constraints upon possible solutions to perceptual problems. This discipline could help theorists to avoid adopting algorithms simply because they work. If Marr's rigorous approach to perception is widely adopted, then some of the mistakes of the past will be avoided. *Vision* is a landmark in the history of perceptual theory.

Endnotes

- The reader should now acquire a copy of Marr's (1982) book, *Vision*. This is quite hard going in places but it is hoped that this chapter will have given the reader confidence to tackle Marr in the original: there really is no substitute.

- Frisby (1979) is an enjoyable, clear, and extremely well illustrated exposition of some of the themes within the computational approach. Written by an obvious admirer of Marr, it conveys much of the enthusiasm of workers in this field.

- Bruce, Green, and Georgeson (1996) is an exceptionally good textbook on visual perception. Although the sections on the computational approach to vision are much more detailed (and difficult) than those in the present chapter, they are well worth working through.

- For a critical but balanced evaluation of *Vision* (Marr, 1982), read Morgan's excellent review (1984).

- Shams and von der Malsburg (2002) list a large number of papers on the behaviour of complex cortical cells.

- Marr's obituary appeared in *The Times* in December 1980.

8 Some final remarks on theories of visual perception

In Chapter 1 we claimed that the contribution of new techniques or methodologies to perceptual research and theory was very considerable. It is hoped that readers will now agree with this claim, after having read about the wide range of techniques reviewed in this volume. Laser scans of the environment, habituation techniques with infants, virtual reality rooms, analyses of optical flow, the creation of pi numbers, fMRI scans, connectionist modelling, spatial frequency analysis, image processing by Gaussian and Gabor filters – each of these has contributed importantly to the work described in the relevant chapters.

Evaluation of approaches to theorizing in perception

Six very different approaches to theorizing in perception have been discussed. At this point a brief evaluative summary will be given of each approach. This will be followed by a general discussion of the problems facing theorists in the general area of visual perception.

The Gestalt theory (Chapter 2) was based upon the numerous discoveries made by proponents of this approach. The Gestalt theorists believed strongly in the dynamic nature of perceiving and the tendency for perception to tend towards coherent, meaningful and simple solutions. The Gestalt demonstrations of the emergent properties of stimulus interactions present an important challenge to all future theories of visual perception. The decision of the Gestalt theorists to concentrate upon strong, reliable effects may provide a lesson for others who wish to make discoveries about perceptual systems. Finally, the emphasis upon phenomenological aspects of perception, which was such an important part of the Gestalt approach, is something that continues to stimulate debate among contemporary theorists: what is it that theories of perception are trying to explain?

The weaknesses of the Gestalt movement lie mainly in the naive approach to theory and explanation. As was shown in Chapter 2, the Gestalt theorists sometimes fell into the trap of mistaking description for explanation. Gestalt theory was, for the most part, not predictive. And when the Gestalt theorists attempted some sort of explanation of the effects they had discovered, they made an unfortunate decision in their choice of a brain model and an equally serious mistake over the selected level of explanation. We attempted to show in Chapter 2 how a modern development in mathematics, algorithmic information theory, might provide a valuable way of quantifying the centrally important Gestalt concept of Prägnanz. This would be an important development in perceptual theory.

Brunswik's probabilistic functionalism (Chapter 3) properly drew attention to a number of hitherto neglected aspects of visual perception: that the cues upon which organisms depend are not certain but only probabilistic; that much of behaviour reveals vicarious functioning; that a careful analysis of the environment from a functionalist viewpoint can sometimes suggest answers to apparently intractable problems. Brunswik's arguments against the classical reductionist approach to experimentation are still relevant and, while his own suggestions concerning the correlational analysis of representatively designed experiments have not been widely adopted, we have described the exciting new techniques and analyses introduced by Purves, Lotto, and their colleagues that seem set to vindicate one of Brunswik's central claims regarding the relationship between distal and proximal stimuli. Tribute must also be paid to Brunswik for his important role in what has been described as 'the inference revolution'.

The weaknesses of Brunswik's approach are, first, that he did not give due recognition to the gains which have been made using the orthodox classical methods which he attacked, methods which uncovered phenomena quite as important and interesting as those he described. Second, the disappointing outcome to some of Brunswik's own experiments suggest that the superior, correlational/multivariate approach he claimed to have designed may not be as easy to apply as he believed: Brunswik did not appear to learn from his own failures. Finally, Brunswik's many stimulating ideas were not communicated in a manner guaranteed to cause others to give them serious consideration.

The neurophysiological approach (Chapter 4) has demonstrated the benefits of combining disciplines. Discovering the neural mechanisms underlying certain perceptual phenomena has been an impressive achievement, one that has confirmed the essential correctness of a number of psychological theories. Knowledge of actual mechanisms has helped some theorists in their work; this knowledge has also provided a useful constraint upon subsequent speculation. Generally, the work described in Chapter 4 is of the highest scientific calibre. One major development, reviewed in Chapter 4, that holds particular promise is the development of fMRI scanning. Now, for the first time, the activity of various regions of the brain can be monitored as conscious volunteers look at controlled stimuli or carry out cognitive tasks. It seems likely that we will witness a torrent of new findings regarding perceptual (and other) functions in the very near future.

In the past, an important weakness of the neurophysiological approach to vision has been its tendency towards reductionism. Another weakness is that the language used inevitably remains ‘within’ the organism. Connectionist network models may overcome the first of these drawbacks. The second is more serious, as it means that the neurophysiological approach cannot, of itself, pay proper consideration to the nature of the environment from which stimuli arise – it would be difficult for it to deal with the probabilistic nature of stimuli, for example. Further, explanations at the level of neurophysiology cannot deal with the subjective nature of seeing – with the phenomenological experiences that reveal the existence of perceptual phenomena in the first place. Finally, knowing that neural systems have certain properties, revealed during experimental research, does not mean that the usual functions of these systems have been discovered. As Marr and others have argued eloquently, knowing *that* a neural system does something does not tell us *why* it does it.

Empiricism (Chapter 5); a good case can be made for the claim that this has been the most successful approach to the formation of a general theory of perception to date. The contents of almost any general text on perception, or any lecture course, comprise in large part the data, explanations and problems unearthed by workers in the empiricist tradition. Empiricism has dominated experimental psychology for a century. A large part of its success lies in the number of powerful demonstrations that are available to shake one’s confidence in the veracity of one’s perceptions: illusions, distorted rooms, context and learning effects, and impossible figures. In fact, we used this technique in Chapter 5 in order to show the reader just how convincing this approach to perception can be.

The doubts about the empiricist approach were described in Chapter 6. Is perception always a constructive process? Are stimuli (or sensations) really so impoverished that the information associated with them needs to be supplemented by memory, reasoning, and so on? Do the problems studied under simplified laboratory conditions adequately reflect the situation facing perceivers in the real world? Does perception proceed essentially in stages? Can the dualism between the organism and the objective world be defended?

At this point, we remind the reader that human perception occurs in two distinguishable environments. The natural environment, in which perception evolved, comprises surfaces and textures, solid objects, rich patterns of multisensory stimulation, movement and change, and so on. But another environment has formed, the age of which is but a moment in evolutionary terms: that of human culture. Here, we have language and symbols, two-dimensional patterns representing three-dimensional things, machines that move us passively through space. It is not surprising that human perceivers can usually cope with this artificial environment: they created it. But the ways in which perception engages with the artefacts of our culture may differ importantly from the ways in which it deals with the natural world. (Further material relevant to this point is included towards the end of Chapter 6, where we describe Norman’s attempt to use recent neurological findings to reconcile the constructivist approach and the position adopted by proponents of direct perception.)

The theory of direct perception (Chapter 6) arose in part as a reaction against empiricism. One of the chief merits of this relatively new approach is the emphasis it places upon the study of the natural environment and the richness of stimulation available to active perceivers. Another merit has been the attempt to counter the distinction between the organism and its ecology, between what happens 'inside' and 'outside' the perceiver. There is a novelty about the direct perception approach, which will sharpen conventional thinking and may ultimately force some major revisions upon constructivist theories. As we attempted to show towards the end of Chapter 6, results from research into the ways in which perceivers achieve 'fits' between themselves and the environment (pi numbers, gaits, intrinsic scaling, and so on) are showing that the theory of direct perception can generate fruitful and testable ideas. For example, optic flow is also being subjected to very precise quantitative analysis in recent research.

The theory of direct perception has its weaknesses. It shows a tendency to underestimate the challenge posed to the visual system by, for example, the need to extract invariants. Some problems have been simply defined out of existence. The resonance model has not yet achieved plausibility, although connectionist networks may eventually suggest solutions to this problem. Finally, the marked differences between this approach and more traditional theories of perception are becoming blurred, as direct perception theorists turn their attention to the various indirect modes of perceiving, which, they accept, are part of human experience. Finally, we describe Norman's attempt to reconcile the empiricist and direct perception approaches by considering the implications of recent neurophysiological work on what are known as the dorsal and ventral pathways within the visual system.

The computational approach to visual perception (Chapter 7) has produced theories that are at present quite narrow (e.g., the theory of stereopsis outlined in this chapter). These are among the most rigorous theories to have emerged in the study of vision. However, where theories arising from within the artificial intelligence paradigm differ from earlier scientific accounts is in the fact that their success points the way towards even more general accounts. To develop this point: consider the successful emergence of a scientific account of colour vision (described in Chapter 4). The combination of evidence from psychology and neurophysiology led to an explanation of both the trichromatic and opponent process aspects of colour perception. But this successful explanation does not of itself suggest how to tackle, for example, the problems of shape or movement perception. In contrast, the success of one application of Marr's approach strongly encourages the belief that the computational approach may be equally powerful when applied to other, very different phenomena.

There are two reasons why Marr attained a high level of rigour in his work. First is the clear distinction he drew between the appropriate levels at which a process may be understood: the computational theory, the algorithm and the implementation. Marr argued, convincingly, that much confusion in perceptual theories in the past arose because of misunderstandings about which level of explanation was appropriate in a particular context. However, there are those who believe that Marr's distinctions between the three levels of explanation are not always appropriate. For example:

> While I do not disagree with Marr's basic argument, I would like to suggest that when dealing with *biological* systems the only sure way to progress is to deliberately get 'confused' between these different levels of analysis. The reason for this is that the organization of biological systems is dictated just as much by constraints of hardware (and by the organism's evolutionary history) as by the 'computational problem'.
>
> (Ramachandran, 1990)

Ramachandran supports this assertion by reminding us that knowledge of the double-helix structure of DNA preceded understanding of its function. Although many modern workers appear to have accepted the value of Marr's distinctions between levels of understanding, Ramachandran's argument appears to carry some weight.

Another reason why the computational approach has achieved such rigour arises from the ways in which theories advanced by Marr and other workers have been rendered explicit. To see whether an idea actually works when written into a computer program is a powerful check against vagueness and imprecision: it is no longer possible to define problems out of existence, neither is it acceptable to explain things by appealing to concepts that (a) remain undefined, such as Prägnanz, or (b) are descriptions rather than explanations, such as 'perceptual constancy'. Everything must be explicit.

One of the arguments against the computational approach has been that it represents a new mentalism; that the computer is an inappropriate model of the perceiver; and that by omitting the phenomenological aspects of perception the theory cannot ever do full justice to its subject matter.

To conclude this section, it can be asserted that there is as yet no satisfactory general theory of visual perception. For example, no theory has adequately united a full analysis of the environment *and* the cognitive aspects of seeing. No general theory has thoroughly incorporated and explained the motor aspects of seeing. The extent to which perception is determined by stimulation (involving bottom-up processes) or knowledge (top-down processes) has not been agreed upon. One theorist (Ramachandran, 1990) finds this state of affairs unsurprising. In arguing for a utilitarian theory of perception, Ramachandran maintains that the perception of, say, colour and motion may have little in common. They comprise part of the 'ragbag' of tricks adopted during evolution. It follows that the search for general laws and general theories of perception may be doomed to failure. Ramachandran's article contains accounts of fascinating discoveries of the extent to which 'quick but dirty' solutions characterize the ways in which vision can be shown to respond when conditions allow. His chapter should now be read in full. The message is a pessimistic one, but it is also thought-provoking.

There is a curiously interesting link between the work of Ramachandran, an active researcher, and a book by the philosopher, Dennett (1991). Dennett's book is a major work and cannot be summarized in a few sentences. For the present purpose, we can refer simply to one of Dennett's examples. Imagine entering a room plastered with small identical photographs of, say, Marilyn Monroe. We look at one and recognize it. Then we become aware of the fact that all walls are covered by this same image. But much of this information is contained in peripheral vision. A question that used to be asked was, how do we fill in or complete our perception under such complicated conditions? Dennett's answer is that we don't: there is no need to. Why not *assume* (albeit unconsciously) that all the pictures are identical? There is no need to work at this – let it be.

Other aspects of visual perception

It has not been possible in an undergraduate text to describe all theories of visual perception. For reasons of space, we have omitted important theories from a number of areas in visual perception. These include: vision and attention, theories of reading, theories and systems of perspective, theories of motion perception, perception and aesthetics, ethology, the effects of brain damage on seeing, cross-cultural studies, and many others. The list is dismayingly long, but interested readers with access to good libraries should easily find enough material to enable them to begin their studies of areas of interest.

Irreconcilable differences

At this point a few general remarks will be offered concerning the theoretical approaches described in previous sections. What, if anything, can be distilled from them?

The first obvious point is that there are too many irreconcilables between the various theorists to permit any general fusion of ideas. The differences, for example, between empiricist views of perception and those of the Gestalt theorists, or between Gibson and Marr, are such that they cannot all be right. Neither does it seem reasonable to suppose that the truth must lie between the rival views: they are too different for that. It is, however, possible to hope that each of the approaches described has merit and that for this reason there may be implications for future theorists. Here are what seem to us to be the best aspects of some of the approaches we have described in this book:

- Laboratory studies of very high quality are now possible. As was shown in Chapter 3, researchers can now take laser samples of the real world and then analyse them in the laboratory. The introduction of virtual reality systems into laboratory settings offers exciting possibilities.
- As stated earlier, new advances in neurophysiology should enable certain new models of the brain to be tested, particularly those of the type proposed in parallel distributed processing models. Neurophysiology will provide knowledge useful for those who work at what Marr described as the stage of 'implementation' of a theory.
- New theories will have to recognize and explain those dynamic aspects of perception discovered by Gestalt theorists, in particular the tendency towards Prägnanz and the fact that stimulus interactions produce new emergent properties.
- Gibson's work has shown: (a) that perceivers are active, not passive, and that sensory and motor systems should be viewed together as integral components of perception; (b) that light may be (usually is) rich in information and that the study of any perceiver should therefore begin with an exhaustive examination of the ecological niche occupied by the perceiver; (c) that the dualism implicit in the animal/environment dichotomy may actually impede our understanding of the nature of perception.
- The role of central, cognitive-like factors in perception has been most brilliantly demonstrated by workers in the empiricist tradition. The question now is to discover the extent to which all perceiving is like this, or whether there are indeed situations where stimulation can specify objects and events without recourse to additional constructive processes.
- To the extent that future theories of visual perception will be self-consciously scientific, they may be guided by the clarity and power of Marr's analysis of what it takes to understand any process. The standards Marr set for clarity and explicitness should serve as a model for theorists in the future.

Those seem to be the safest and best conclusions arising from any comparison of the selection of theoretical approaches outlined in this book. We shall return to some of these conclusions during the following sections that contain some further general remarks on theorizing.

The remainder of this chapter is in four sections. First, an attempt will be made to show the challenge offered by visual perception; why it can be expected to continue to fascinate researchers and theorists. Then a list will be given of some of the achievements to date: the gains that have been made because some theorists have attempted to develop general theories of the type described in previous chapters. We shall then repeat the claim that there has not yet been a satisfactory theory of vision, and, in another section, will attempt to explain this by describing some of the problems facing theorists in this area. Finally, some speculations will be offered concerning the next generation of theories of visual perception.

The challenge of vision

The theme of this book is seeing and attempts to explain it. But seeing is part of the daily life of anyone who can read this book. We are as familiar with seeing as with anything – so familiar that it is easy to overlook what an achievement seeing represents. Underlying this awareness of a solid, coherent world are the activities of many neurons – saline-filled tubes – and that is all. How do they do it? That is the ultimate question, but it is unlikely to be answered for a very long time. Here are just three examples of particular problems arising in the study of visual perception.

What is meant by 'seeing'?

If we consider once again the case of the patient S.B., who recovered his sight after long periods of blindness, it will be remembered that he could not recognize a lathe until he touched it: 'Now that I've felt it I can see' (Gregory & Wallace, 1963).

What did S.B. mean by 'see'? The visual image of the lathe remained the same, but something must have changed in S.B's head. What was it? In what way could he *not* see before using his sense of touch? Was the change in S.B's visual perception akin to Fodor and Pylyshyn's (1981) distinction between merely seeing the Pole star and seeing it *as* the Pole star? These are surely questions for a general theory of visual perception.

Different eyes

Reference has been made throughout this book to 'the eye' and 'the visual system'. In fact, apart from a few references to vision in monkeys and cats, all the main discussions have been about the human eye and the human visual system. But there are millions of different eyes. Vision in insects, for example, is based upon a different set of structures from those of any vertebrate, and visual information goes to a very different form of central nervous system. At present, we can only make a few educated guesses at what it must be like to see as a fish, a bird, or a flea. Even the lives of our pets remain deeply mysterious. It is a serious flaw in recent documentary broadcasts on animal vision that visual input is mistaken for visual perception. A familiar trick is to film the world through a segmented lens and then say that this is what the perception of an insect with a compound eye must be like. No – this is a major fallacy: we view the resulting picture and interpret it with highly developed brains; it is simply impossible to know how the output from a segmented eye looks to an insect.

Building seeing machines

There is one activity that makes visual research different from that in other sensory areas, and almost unique in science generally. One of the goals of a growing number of researchers is to understand vision and then build a seeing machine. To date, this has inspired some highly original and interesting work. And it has also provided a valuable check upon loose speculation and the imprecise analysis of problems.

> . . . The first great revelation was that the problems are difficult. Of course, these days this fact is a commonplace. But in the 1960s almost no one realized that machine vision was difficult. The field had to go through the same experience as the machine translation field did in the fiascos of the 1950s before it was at last realized that here were some problems that had to be taken seriously. The reason for this misperception is that we humans are ourselves so good at vision.
>
> (Marr, 1982)

In other words, building a seeing machine is going to be very difficult, not least because there are some formidable philosophical problems associated with this whole enterprise which raise doubts as to its feasibility: for example, is there any sense in which a seeing machine would be conscious? For now, it is sufficient to say that here is a real goal for general theorists. But if it can be attained and a machine becomes able to do enough to make us say that it can see, then the builders may be in the unique position of knowing with certainty that at least some of their ideas on vision are correct. The reader will not be surprised to learn that this is a research area of vast commercial potential. Robots able to discriminate quickly and accurately between objects in the environment would have immense value. The implications of machine vision are so important that this is one area where researchers are unlikely ever to be short of funds.

Of course, no existing theory has come close to meeting any of these challenges. They were included at the start simply to show what fascinating challenges they are, fascinating enough, surely, to continue to attract ingenious and creative people into visual research.

Some of the gains made during the search for general theories

Here are some reasons to be optimistic over the future of research into visual perception:

- *Many general properties of vision have been discovered.* For example, it is one thing to know that observers tend to see objects as having constant size irrespective of distance, and to be able to measure the magnitude of this effect. It became even more interesting when it was realized that shape perception also tends to veridicality. But when the same effect became apparent in colour perception, it was clear that here was a general characteristic of visual perception. (The realization that a comparable effect can occur with loudness suggests that the effect is even more general.) However, to recognize the fact that perceptual constancy *is* a general characteristic of vision requires a theory, for data from different experiments are merely data: something more is required before their more general significance can be recognized, and that something is a theory (in this case it was the Gestalt theory).

All the theories described in this book have become broader with time. They have sought to embrace increasingly disparate sets of phenomena. For example, Gibson's demonstrations of the role of invariants in the perception of simple geometric shapes has been extended to include the match between the cardioidal strain transformation and the ageing of the human face. Even further from the original demonstration is the discovery of invariants in sound patterns that can specify the likelihood of collision. But would sound patterns have been examined for this property prior to Gibson's theory?

- *The range of application of theories has broadened.* For example, the empiricist/constructivist approach convinced many that here was a theory that could account for many of the dynamic aspects of perception, such as the ability of perceivers to go beyond the sensory evidence – the closing of gaps, the correct identification of ambiguous stimuli through the use of context, and many other tendencies which have been described in Chapter 5. But then an extrapolation of the same theory to the geometric illusions allowed Gregory to suggest 'inappropriate constancy scaling' and Day to invoke 'perceptual compromise' as possible explanations of these strange phenomena.
- *Theories have inspired improvements in methodology.* When Purves, Lotto, and their colleagues began to develop their empirical theory of perception, they needed some way of accurately sampling the environment and comparing the distal and proximal stimuli arising from it. Their solution, outlined in Chapter 3, was to develop methods of analysing inputs from computer scans of the environment. Having done that, they have introduced into visual science a new tool, and one of staggering potential – in the present author's view.

In similar manner, algorithmic information theory, an abstract and specialized branch of mathematics, has now provided workers in vision with an objective measure of simplicity – a measure that can finally test certain important claims made by the Gestalt theorists.

Of course, we have repeatedly made a claim in the other direction, namely that developments in technique exert important influences on perceptual research and theory, but, as we have shown above, this remains a two-way process.

- *Theories have become more precise and scientific.* Two of the hallmarks of any scientific theory are that its terms are precisely defined and that the theory is testable and therefore falsifiable. Of course, not everyone agrees that the best solution to problems in psychology will always be scientific ones, a point that will be returned to later. For the moment we shall assume that it is in fact desirable for theories of perception to be as scientific as possible.

  If this assumption is agreed upon, then it can be claimed that there has been real progress in theorizing. Consider again the above example referring to the Gestalt theory (Chapter 2). Certain key terms within the theory lacked precise operational definition. For example, the concept of Prägnanz was described in a manner that made it seem very interesting and important, with many convincing illustrations. But how did one measure it? How could one be sure that some novel stimulus array would show high or low Prägnanz? Was Prägnanz a descriptive shorthand for such factors as balance, simplicity, and symmetry, or was it an explanatory concept: things being seen as coherent wholes *because* they had high Prägnanz? Using Kolmogorov complexity measures, it should be possible in the near future to answer this question.

  Brunswik wrote of the ‘stupidity of the senses’, by which he meant the inability to compensate for visual illusions once one knows the details of their construction. But to what extent did he claim this as a general principle? There are, after all, numerous examples (some of which would have been known to Brunswik) showing that perceivers can be very flexible in their behaviour: the recovery from mild sensory distortion, the ability to reinterpret ambiguous patterns, or to extract meaning from corrupted or noisy displays. Where does ‘stupidity’ begin and end?

  In contrast, more modern theories of visual perception are much more precise. Gibson made important use of the term ‘optic flow’ in his theory. More recent work has shown possible constraints upon what optic flow can in fact specify (it doesn’t indicate the point of collision if a surface is approached with the head slightly averted, for example). And Gibson seriously underestimated the difficulty of extracting the higher-order invariants from optic flow, as Marr was able to show. Nevertheless, there was never any doubt over what Gibson actually meant by the term ‘optic flow’; it was well defined.

  By the time Marr’s computational theory of vision was published, even higher standards of precision had been reached. One of Marr’s enduring contributions will probably be the rigour that he introduced into the activity of theory construction. We have already described Marr’s distinctions between the computational theory, the algorithm, and the implementation. But the whole of his book *Vision* is an object lesson in precision and clarity, attributes that owe much to the discipline required when converting ideas into working computer programs.

These examples serve to support the claim that real progress has been made. Nevertheless, formidable obstacles remain.

Problems remaining

First, it must be accepted that vision must be very complicated indeed. Recognizing this fact will make the next section seem less negative and pessimistic. Think of the things that vision can do. Reflect on the last time you drove or were driven down a motorway on a wet night. That you are alive to tell the tale says a lot about vision: it provides the basis for judgements of speed and distance under difficult conditions, for moving in and out of lanes, steering half a ton of steel between and around hazards. All this while thinking about the depth induced by random-dot stereograms. Or think of being forced to sit at the end of the front row in a crowded cinema: how distorted the screen seems at first, but how quickly it comes to appear normal. Then think about walking down the cinema steps without consciously looking at them.

All the phenomena described in this book reinforce the conclusion that vision is remarkable. It has defeated some of the best thinkers of their time. There have been few better scientists than Helmholtz, and he left many problems unsolved. The Gestalt theorists were clever and creative researchers, but their theory is flawed. Marr has been described as a genius, and yet some aspects of his work have been shown to be wrong.

Three of the formidable problems facing any who search for general theories of visual perception will now be described.

The definition of a stimulus

It is manifestly true that the job of vision is to inform us of things and events in the external world. The medium by which information is carried to the eye is, of course, light. Sometimes light is informative, sometimes not. When it is informative, the term ‘stimulus’ tends to be applied. It is therefore clearly desirable, when constructing theories, to be able to define, measure, and, where necessary, control stimuli. But what *is* a stimulus?

This is one of those problems that get harder the more one thinks about them. And there is a noticeable lack of agreement between theorists over this basic issue. It is also one that has exercised many philosophers. A sensible aim would be to try to describe any stimulus in terms of its physical characteristics – this would seem to be a natural starting point for any scientific endeavour. After all, part of the success of chemistry and physics, at least in the early days of those subjects, lay in the ability to define and measure such basic entities as atoms and molecules. Is it likely that a comparable degree of objectivity and precision over stimuli will eventually be achieved? The answer is, probably not. Here are some reasons why.

Imagine that one is participating in a psychophysiological experiment.

Electrodes have been attached to the back of the head in order to monitor some of the activity of the visual cortex. Every second a flash of light is delivered to the eye. Then, without warning, the sequence is broken by omitting one of the flashes. What happens? Well, the gap will be noticeable – it will capture the attention as a novel event. And the subsequent increase in activity in the visual cortex will confirm this introspection. Now, is a missing flash a stimulus? It surely possesses many of the characteristics we would wish to assign to stimuli on commonsense grounds. But can it be measured? No, of course not, for the missing light has no physical existence: it is something that might have occurred but did not – the dog that didn't bark in the night.

An equally important phenomenon has been described in the account of Gestalt theory (Chapter 2): sources of stimulation interact to yield novel perceptions. In the phi phenomenon, for example, the seen movement cannot be explained or predicted from a description of either of the pair of inducing lights – it depends upon their spatial and temporal relationships. At the heart of the Gestalt theory is the axiom that wholes are more than the sum of their parts, and Gestalt publications bristle with demonstrations of this effect, although the Gestalt theorists were never able to offer quantitative data on this point.

Thus, it seems certain that any description of a stimulus in isolation will prove to be inadequate and that it will usually be necessary to consider other stimuli which (a) might have been present, or (b) are present and capable of interacting with the original stimulus.

Another difficulty arises from the fact that stimulus definition is theory-dependent. Visual scientists would not describe patterns in terms of their Fourier transforms unless they believed that the visual system also performed such analyses. If it is assumed that perceptions are hypotheses, then a stimulus is something that can stand as evidence against which these can be tested. In holding to this view, one is clearly showing a degree of theoretical bias. Similarly, Gibson objected to any definition of a stimulus as a momentary happening, something frozen in time. Instead, he considered that that which endured or was invariant over change was the basis for a given percept. Gibson also emphasized that activity on the part of the perceiver was a source of stimulation, although it is not always easy to see how such stimulation could be measured. Marr's focus is upon spatial variations of intensity and so on in the visual image. Although he acknowledges that Gestalt-type interactions occur, his theory nevertheless concentrates upon the analysis and processing of separate parts of the image: none of his algorithms deals specifically with whole/part effects, neither is it easy to see how any could. The differences between these theoretical approaches go some way towards explaining why there is no single definition of a stimulus.

The appropriate level of explanation

At issue here is the level at which visual phenomena should be described and explained.

It was pointed out in Chapter 4 that there have always been those who prefer to explain perceptual phenomena in terms of neural mechanisms, rather than psychological constructs. This form of reductionism has yielded important insights into the nature of perception, and the interaction between the disciplines of experimental psychology and neurophysiology has been a fruitful one. But, as was argued in Chapter 4, there is a fundamental flaw in the idea that the eventual explanation of perception will be physiological; namely that neurophysiology remains 'inside' the organism, while perception involves the external world. Neural events may be isolated entities, but stimuli arise from within a context, a context that shapes our conscious experience. A general theory of vision, should such a thing ever be possible, would have to respect this fact. Thus, the language of such a theory would be more likely to be psychological than physiological.

However, even if the emphasis in the future is away from physiological reductionism, there will still be a problem over the best level at which to write a theory. Assumptions of similarity between humans and computers led to computational theories. Recognition of certain dynamic properties of vision encouraged thinking in terms of vectors and fields. Those who claim that phenomenal experience must be accounted for by theories will write in the appropriate language. And anyone wishing to theorize only at the most formal and abstract level will adopt the language of mathematics. It is a sobering thought that, years from now, a general theory of vision may indeed be so abstract and complex that few workers now alive would be able to understand it. But it is as yet unclear what level of explanation will be adopted in such a theory.

The place of subjective experience in perceptual theory

Suppose that there was a respected general theory of visual perception, a theory that successfully handled the main phenomena of colour, shape, depth, motion, and so on. Would the theory be limited to human perception? Many theorists have clearly believed that their work could embrace non-human species: the Gestalt theorists demonstrated some of their effects in apes and chickens; Brunswik studied probability matching in rats; direct perception theorists talk about different species in different ecological niches; Marr speculates about the landing behaviour of the house fly; cortical edge detectors were first demonstrated in cats. However, although much is known about vision in a few non-human species, we will never fully understand the quality of their conscious experiences, if any. As was stated earlier, we cannot know what it is like to see as a fly.

We are, however, intimately aware of our own consciousness, of what seeing is like 'from the inside'. In actual research it is possible to adopt a tough-minded, behaviourist approach to perception, measuring human and animal performance in similar ways. There is then no place for introspection, only controlled, measurable responses. But where do the researchers' ideas come from, if not from their own subjective experiences? The Purkinje shift – changes in the relative brightness of blue objects with changing illumination – can be demonstrated by monitoring the behaviour of humans or animals. But why should anyone think to do so, unless they had first noticed the effect (Purkinje was lying in bed when he noticed the effect – off colour, perhaps?). Should workers in perception deliberately omit from their theories parts of their daily experience?

Discussions such as this may be dismissed as relevant only to the origins of scientific theories, not to the logic of the theories themselves. After all, it could be argued, much has been learned about the behaviour of honeybees without anyone knowing what it is like to use the sun's position in a signalling dance. And one of the most powerful binocular illusions, Pulfrich's pendulum,[1] was discovered by a one-eyed man. But, for example, would purely objective research into colour vision ever have discovered that certain colour combinations are very unpleasant, or that some colours appear warm, others cold, or that some people hear coloured sounds?

It is equally important that experience also tells us what does *not* happen during perception. For instance, our noses are constantly in view but not seen: no formal account of a visual experiment ever includes the nose in a description of the stimulus array – it doesn't need to, for noses are not important in seeing, and it is subjective experience that tells us this. We can learn to see that objects occupy less of the visual field when they recede from us (and artists must be able to do this), but phenomenally they remain the same size. Which of these two ways of perceiving the objects should the theorist be concerned with?

The strongest claims for the inclusion of subjective experience in accounts of perception have been made by modern phenomenologists. They insist that our perception of, say, a house, transcends any limited vantage point: we 'see' the volume of the house, its solidity, even when the only visible aspect is the front. Our phenomenal experience includes the knowledge that we are 'inside' our bodies. We know what things would look like from alternative vantage points.

The problem of conscious experience has now been alluded to several times in this book. It is, however, so important that one final example must be presented.

---

[1] When a pendulum, swinging across the visual field, is viewed with a dark filter in front of one eye, the swing changes into a 3D ellipse in which the pendulum appears to move towards and away from the viewer.

We have all had the experience of perceiving some thing or event with unusual clarity. At its most dramatic, this happens when we witness something sudden and violent, such as an accident. But there are other times when one simply feels calm or quiet and yet strangely attentive. It is as if one has achieved a new sharpness of focus in which things are seen almost as if for the first time.

Developing a scientific account of seeing is very different from the creation of an original work of art; the two activities attract different personalities motivated by different goals. But when it comes to describing the more elusive aspects of experience, who is to say that the artist is not the better analyst? Here is a short sequence from a twentieth-century novel. After reading it, do you not agree that it is reminding us of what it is like to perceive something intensely; that the description is as interesting in its way as anything to be found in more scientific accounts?

Two men are making a coffin:

> The lantern sits on a stump. Rusted, grease-fouled, its cracked chimney smeared on one side with a soaring smudge of soot, it sheds a feeble and sultry glare upon the trestles and the boards and the adjacent earth. Upon the dark ground the chips look like random smears of soft pale paint on a black canvas. The boards look like long smooth tatters torn from the flat darkness and turned backside out.
>
> Cash labours about the trestles, moving back and forth, lifting and placing the planks with long clattering reverberations in the dead air as though he were lifting and dropping them at the bottom of an invisible well, the sounds ceasing without departing, as if any movement might dislodge them from the immediate air in reverberant repetition. He saws again, his elbow flashing slowly, a thin thread of fire running along the edge of the saw, lost and recovered at the top and bottom of each stroke in unbroken elongation, so that the saw appears to be six feet long, into and out of pa's shabby and aimless silhouette . . .
>
> The air smells like sulphur. Upon the impalpable plane of it their shadows fall as upon a wall, as though like sound they had not gone very far away in falling but had merely congealed for a moment, immediate and musing . . .
>
> (William Faulkner, *As I Lay Dying*)

Faulkner has used his formidable powers of description to put into words something of the essence of perceiving. In a sense, we are there with him in the Deep South, watching this strange scene. And the ways in which he has noticed things – the muffled sounds, the subtle gradations of light and shade – remind us that there is a lot to the business of perceiving, and a long way to go before it is understood. Will computer simulations ever do justice to these aspects of awareness? Can they be captured in scientific accounts of perception? It will be fascinating to see how these formidable problems are approached in the future.

The evolutionary background to human vision

About 6 million years ago the earliest humanoids split from the apes. Within another million years our ancestors had become fully bipedal. Then, in the last 2 million years, the human brain developed more rapidly than any other organ in evolutionary history. It was during these last 2 million years that the human visual system attained its present form and, presumably, its remarkable range of functions.

But the environment that exerted the selective pressures that shaped the evolution of perception was very different from that which most of us inhabit today. Now our daily lives bring us into regular contact with signs and symbols. Most of us live in a built environment – a place of sharp edges, flat surfaces and artificial lighting. Our bodies have a natural speed across two-dimensional surfaces of 4 or 5 miles per hour, but are frequently moved passively through three dimensions at speeds 100 times greater than this. There are many highly unnatural environments.

Although this question has been raised several times in this book, it is worth one last mention: can we expect a single theory to explain perception of the natural *and* the artificial world? That is to say, there may be one set of mechanisms (describable by one set of laws) which will ensure that, for example, singleness of vision and stereoscopic fusion are achieved in the lit environment; but there can be no such inbuilt mechanism to allow us to fly in clouds. That is something that we can learn to do, provided we have access to the right instruments. Instrument flying undoubtedly involves perception as well as skill, but is it the same sort of perception as, say, normal stereopsis? If not, where is the dividing line to be drawn?

As has been shown in previous chapters, the extent to which a perceptual theorist remains aware of the evolutionary background to vision determines, in part, those things which are emphasized in his or her theory. R. L. Gregory is much more concerned to explain illusions than is J. J. Gibson, who regards these as arising from essentially unnatural patterns of stimulation. It is almost a truism that the more a theorist favours a cognitive approach, the greater will be the emphasis upon those things over which we cognate: pictures, puzzles, words, lists. But is seeing a word the same as seeing a face? Do we 'read' these two types of pattern in the same manner? This leads naturally to a much larger question, as to which sets of phenomena – the natural or the artificial and symbolic – should be researched. Which offers the greater chance of success? This is clearly an issue over which there will continue to be much controversy.

Theories and the future

Futurology is not to everyone's taste. It is certainly not this author's. And yet the reader who has come this far has a right to some sort of personal statement; after such a lengthy attempt to be dispassionate and fair, it seems only right to offer some more personal views on theories of visual perception. This should not be read as a prediction concerning 'the' final theory of vision: indeed, there will probably never be one – rather, it is a guess concerning the next important general accounts of visual perception, the future equivalents of something as important in their way as was the Gestalt theory. Here are eight assertions. They should be read with an appropriate degree of scepticism:

- *Rivalry between different theories will continue.* This claim is reinforced by many of the points listed below. It is going to be very difficult, if not impossible, to reconcile a scientific account of perception (possibly based on a computer model) with the claims of phenomenologists. The need to describe and explain human awareness can be expected to form the ground for much future debate among theorists.
- *The phenomenological component in theories will grow.* This is predicted simply on the basis of a probable future swing away from the reductionist, mechanist approach that has dominated Western psychology this century.
- *General theories will be mainly concerned with human perception.* It will be possible to explain some aspects of perception in other species, of course. Much is already known about, say, vision in insects. But most species are simply too different from us in structure and life-style for their perceptions ever to be predicted from a human theory. We know how bats navigate, but cannot imagine their phenomenal world.
- *Future theories will include more thorough analyses of the environment.* Although many workers have respected the need to understand the environment to which the visual system responds, there have been relatively few experimental studies of this environment. One reason for this is that the best tools available to vision researchers were confined to laboratories. Apart from cameras and similar instruments, the best had to be left behind indoors. However, we have seen the start of what looks to be a highly important development in Chapter 2, when the empirical theory of vision was outlined. A large part of our account of this theory was devoted to a new technique allowing the environment to be scanned with an unprecedented degree of accuracy. The success of this research will surely inspire others to start similar investigations of environmental variables.
- *Theories will have a functional bias.* To repeat Jung's phrase, 'We are of an immense age'. As an increasing effort is made to plot the evolutionary background from which humans emerged, rapid gains in

knowledge seem almost inevitable. Recent advances in evolutionary theory have explained such apparently baffling problems as the Panda's sixth finger (read Gould, 1980, in order to experience the excitement of this type of research). Can we not expect similar researches to uncover the functional significance of, for example, the human tendency to see vertical lines as longer than physically equal horizontal ones; why faces with large pupils strike us as more attractive; or why the human visual system is modifiable mainly during the first few months of life?

- *Research and theory will continue to be influenced by technical developments.* Those who have not experienced the effects of some major new technique in perceptual research may find it hard to believe what an impact these can have on ways of thinking. For example, it was customary until recently to describe the performance of the eye in terms of (a) its acuity, (b) its sensitivity. The advent of visual gratings and the associated idea of spatial frequency analysis, tuned spatial frequency channels, and the contrast sensitivity function changed our ideas about these aspects of vision completely. This is a very familiar story in psychological research and could be illustrated by many more examples. The point is that psychologists and neurophysiologists are not the only people who are thinking about problems of seeing: engineers, physicists, zoologists, and mathematicians can all claim a legitimate interest in the problem. As more and more of their techniques are adopted, the impact of these other disciplines on perceptual theory and research can be expected to become increasingly dramatic.
- *Models will become increasingly important.* This relates to the last point and is too obvious to need stressing. We have already shown how available models determine psychological thinking. As new and more powerful computer systems emerge, these will inevitably be used as models of the brain by many psychologists.
- *Simulation will become increasingly common.* Thirty years ago very few workers in perception had access to computers; now they are to be found in every laboratory. It would be hard to exaggerate the impact of these remarkable machines in experimental psychology and neurophysiology. That they have provided a very seductive model of the perceiver has been discussed at some length in earlier chapters. Computers have also changed the ways in which actual experiments are conducted. For example, it is no longer necessary always to specify the ordering of stimulus presentations at the start of an experiment. Because the computer can react much faster than any experimenter, the observer's performance can be assessed from trial to trial, and subsequent presentations can be modified accordingly. As an example of the possibilities afforded by this speed and power, we may simply cite modern eye-movement studies, in which stimulus presentations can be given at precise intervals *following* the initiation of eye movements. Such work

> would have been impossible prior to the advent of computer-controlled displays.
>
> There is, however, another way in which computers are having an impact on perceptual research and theorizing, and this is their ability to simulate processes. In the past, a typical sequence would be along these lines: someone has an idea – a possible explanation of some perceptual phenomenon, a critical test of a controversial theory, or simply a hunch about how something comes about. An experiment is designed, equipment built, participants recruited. Some weeks or months later the data are ready for analysis. Interesting results may or may not be found and the experiment may be repeated with certain modifications, and so on. This is slow, inefficient, and often frustrating work – ask any experimenter. It is often realized, half-way through an experiment, that something is wrong, that things could have been done in a different way. However, the rules of experimentation and the demands of statistical analysis mean that the experiment must be run to the end. These are the major problems: the minor ones are that people don't turn up, mothers of young infants change their minds, and animal research is discouraged.
>
> But some ideas about perception can be tested, at least in the first place, without using observers at all: a computer will suffice. What is required, of course, is complete clarity of thought. Each and every part of the hypothetical process to be simulated must be programmed. But this can be a valuable discipline: it is no longer possible to be vague, to employ undefined terms; everything must be made explicit for the computer. Once started, the simulation will be run at high speed, and the input variables can be modified at will. In this way, several thousand hours of possible experimentation can be compressed into a few days. We have described one famous use of simulation to test a perceptual hypothesis in Chapter 7, when the Marr–Poggio model of stereopsis was outlined. The success of this and other recent simulations is a clear pointer to the future.

That is the last speculation. One thing seems certain, however, and that is that vision will always fascinate. The visual sense is, above all, the channel through which curiosity becomes manifest. Think, for example, of the crowds that gather to see the rare artefacts of other civilizations: Tutankhamen's mask, the horses of ancient China. Watch people at sporting events, in the cinema, in front of television sets. Ask why it is that every well-known beauty spot has a continual stream of visitors, gazing out across the view.

There has as yet been no satisfactory general theory to explain how we see the world, none that has been able to satisfy all demands of breadth, precision, and falsifiability that are required of good theories. It seems to the present author that in fact there will never be a truly comprehensive theory of visual perception. Too many approaches are possible; too many levels of explanation exist; different theorists ask different questions; no single language – neurophysiological, psychological, phenomenological, or mathematical – can hope to cover all aspects of seeing. We should not be disheartened: visual perception utilizes not only the eye – which is a structure of formidable complexity – but the brain, which in humans comprises ten thousand million cells interacting in ways as yet not understood. Underlying our experience of seeing is the most complicated system ever known. However, by scrutinizing each new theory as it appears, looking for its strengths and weaknesses and attempting to compare it with rival theories, our understanding must grow. There may never be a general theory of visual perception, but there are many interesting challenges ahead.

Endnotes

- Some of the issues addressed in this chapter are clearly philosophical. Readers wishing to know more about theory, explanation, and the philosophical problems associated with reductionism and the computer metaphor of the mind should start by reading the following: Chalmers (1982), Dreyfus (1972), and Russell (1984).

- For somewhat more advanced discussions of artificial intelligence and the computer metaphor, see Winograd and Flores (1986).

- Contemporary editions of the journal *The Behavioral and Brain Sciences* should be watched for interesting debates on some of the issues raised in this chapter.

References

Adrian, E. D. (1928). *The basis of sensation*. London: Christophers.

Ames, A. (1949). *The nature and origins of perceptions* (preliminary laboratory manual). Hanover, NH: The Hanover Institute.

Arnheim, R. (1949). The Gestalt theory of expression. *Psychological Review*, *56*, 156–171.

Arnheim, R. (1956). *Art and visual perception*. London: Faber.

Arnheim, R. (1969). *Visual thinking*. Berkeley, CA: University of California Press.

Arnheim, R. (1987). Prägnanz and its discontents. *Gestalt Theory*, *9*, 102–107.

Atkinson, J., Hood, B. M., Wattam-Bell, J., Auker, S., & Trickelbank, J. (1988). Development of orientation discrimination in infancy. *Perception*, *17*, 587–595.

Attneave, F. (1955). Symmetry, information, and memory for patterns. *American Journal of Psychology*, *68*, 209–222.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology*, *119*, 69–88.

Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Bartlett, M. S., Movellan, T. T., & Sejnowski, T. T. (2002). Face recognition by independent component analysis. *I.E.E.E. Transactions on Neural Networks*, *13*(6), 1450–1464.

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception and Psychophysics*, *1*, 300–302.

Blakemore, C. (1974). Developmental factors in the formation of feature extracting neurons. In F. G. Worden & F. O. Smith (Eds.), *The neurosciences, 3rd study program*. Cambridge, MA: MIT Press.

Blakemore, C. (1990) (Ed.). *Vision: Coding and efficiency*. Cambridge: Cambridge University Press.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system sensitive to the orientation and size of retinal images. *Journal of Physiology*, *203*, 237–260.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science*, *166*, 245–247.

Brehmer, B. (1984). Brunswikian psychology for the 1990s. In K. M. J. Lagerspetz & P. Niemi (Eds.), *Psychology in the 1990s*. Amsterdam: Elsevier: North Holland.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.

Brown, S. C. (Ed.) (1974). *Philosophy of psychology*. London: Macmillan

Bruce, V., Green, P. R., & Georgeson, M. A. (1996). *Visual perception: Physiology, psychology and ecology* (2nd ed.). Hove, UK: Psychology Press.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*, 123–152.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, *42*, 33–44.

Brunswik, E. (1938). Psychology as a science of objective relations. *Philosophy of Science*, *4*, 227–260.

Brunswik, E. (1939). Probability as a determiner of rat behavior. *Journal of Experimental Psychology*, *25*, 175–197.

Brunswik, E. (1948). Statistical separation of perception, thinking and attitudes. *American Psychologist*, *3*, 342.

Brunswik, E. (1952). The conceptual framework of psychology. In *The International encyclopedia of unified science* (Vol. 1, p. 10). Chicago, IL: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of 'proximity' and other Gestalt factors. *American Journal of Psychology*, *66*, 20–32.

Butterworth, G. (1983). Structure of the mind in human infancy. In L. P. Lipsitt & C. K. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 2). Norwood, NJ: Ablex.

Butterworth, G. (1988). Events and encounters in infancy. In A. Slater & G. Bremner (Eds.), *Infant development*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Butterworth, G., & Hicks, L. (1977). Visual proprioception and postural stability in infancy: a developmental study. *Perception*, *6*, 255–262.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–566.

Campos, J. J., Kermoian, R., Witherington, D., Chen, H., & Dong, Q. (1997). Acuity, attention and developmental transitions in infancy. In P. J. Lang & R. F. Simons (Eds.), *Attention and orienting: Sensory and motivational processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Campos, J. J., Berlunthal, B. I., & Kermoian, R. (1992). Early experience and emotional development: the emergence of wariness of heights. *Psychological Science*, *3*, 61–67.

Campos, J. J., Anderson, D. I., Barbu-Ruth, M. A., Hubbard, E. H., Hertentein, H. J. & Witherington, D. (2000). Travel broadens the mind. *Infancy*, *1*(2), 149–219.

Chaitin, G. J. (1966). On the length of programs for computing finite binary sequencies. *Journal of the Association for Computing Machinery*, *13*, 547–569.

Chaitin, G. J. (1990). *Information, randomness and incompleteness* (2nd ed.). Singapore: World Scientific.

Chaitin, G. J. (1999). *The unknowable*. Singapore: Springer-Verlag.

Chalmers, A. F. (1982). *What is this thing called science?* (2nd ed.). London: Open University Press.

Chasles, M. (1830). Note sur les propertiétés generales du système de deux corps semblables entr'eux et placés d'une manière quelconque dans l'espace; et sur le déplacement fini ou infiniment petit d'un corps solide libre. *Bulletin des Sciences Mathématiques (Férussac)*, *14*, 321–326.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566–581.

Coren, S. (1986). An efferent component in the visual perception of direction and extent. *Psychological Review*, *93*, 391–410.

Costall, A. (1981). On how so much information controls so much behaviour: James Gibson's theory of direct perception. In G. Butterworth (Ed.), *Infancy and epistemology*. Brighton, UK: Harvester Press.

Craig, C. M., Delay, D., Grealy, M. A., & Lee, D. N. (2000). Guiding the swing in golf putting. *Nature*, *405*, 295–296.

Creed, R. S., Denny-Brown, D., Sherrington, C. S., Eccles, J. C., & Liddell, E. G. T. (1932). *Reflex activity of the spinal cord*. Oxford: Clarendon Press.

Cutting, J. E. (1982). Blowing in the wind: Perceiving structure in trees and bushes. *Cognition*, *2*, 25–44.

Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.

Cutting, J. E., Vishton, P. M., & Braren, P. A. (1995). How we avoid collisions with stationary and moving obstacles. *Psychological Review*, *102*, 627–651.

Dartnall, H. J. A., Bowmaker, J. K., & Mollon, J. D. (1983). Human visual pigments: microspectroscopic results from the eyes of seven persons. *Proceedings of the Royal Society of London, Series B*, *220*, 115–130.

Dawkins, R. (1989). *The selfish gene* (2nd ed.). Oxford: Oxford University Press.

Dawkins, R. (1996). *Climbing Mount Improbable*. London: Viking.

Day, R. H. (1989). Natural and artificial cues, perceptual compromise and the basis of veridical and illusory perception. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures and mechanisms* (pp. 107–129). Amsterdam: Elsevier Science Publishers.

Day, R. H., & Duffy, F. M. (1988). Illusions of time and extent when the Müller–Lyer figures moves in an aperture. *Perception and Psychophysics*, *44*(3), 205–210.

Day, R. H., & Kasperczyk, R. T. (1985). Apparent displacement of lines and dots in a parallel-line figure: A clue to the basis of the Poggendorff effect. *Perception and Psychophysics*, *38*, 74–80.

Day, R. H., & Power, R. P. (1965). Apparent reversal (oscillation) of rotary motion in depth: An investigation and a general theory. *Psychological Review*, *72*, 117–127.

Dennett, D. C. (1991). *Consciousness explained*. London: Penguin Science.

De Valois, R. L. (1960). Color vision mechanisms in monkey. *Journal of General Physiology*, *43*, 115–128.

Dienes, Z., & McLeod, P. (1993). How to catch a cricket ball. *Perception*, *22*, 1427–1439.

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. New York: Harper & Row.

Duchon, A. P., Warren, W. H., & Kaelbling, L. P. (1995). Ecological robotics: Controlling behaviour with optical flow. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dziurawiec, S., & Ellis, H. D. (1986). Neonates' attention to face-like stimuli: Goren, Sarty and Wu (1975) revisited. Paper presented at the Annual Conference of the Developmental Section of the British Psychological Society, University of Exeter, UK, September 1986.

Earle, D. (1985). Perception of glass pattern structure with stereopsis. *Perception*, *14*, 545–552.

Ehrenfels, von C. (1890). Über Gestaltqualitäten. *Vierteljahresscher für Philosophie*, *14*.

Ellis, W. D. (Ed.) (1938). *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.

Feynman, R. P. (1999). *The Feynman lectures on gravity*. London: Penguin Science.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's 'ecological approach'. *Cognition*, *9*, 139–196.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, *28*, 3–71.

Frisby, J. P. (1979). *Seeing*. Oxford: Oxford University Press.

Ghim, H.-R. (1990). Evidence for perceptual organization in infants: Perception of subjective contours by young infants. *Infant Behaviour and Development*, *13*, 221–248.

Gibson, E. J. (1982). The concept of affordances in development: the renascence of functionalism. In W. A. Collins (Ed.), *The concept of development. The Minnesota symposia on child psychology*, *15*, 55–81.

Gibson, J. J. (1950). *The perception of the visual world*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1961). Ecological optics. *Vision Research*, *1*, 253–262.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1967a). New reasons for realism. *Syntheses*, *17*, 162–172.

Gibson, J. J. (1967b). Autobiography. In E. G. Boring & G. Lindzey (Eds.), *History of psychology in autobiography*. New York: Irvington.

Gibson, J. J. (1971a). A preliminary description and classification of affordances. Unpublished manuscript, reproduced in E. Reed & R. Jones (Eds.), (1982), *Reasons for realism*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibson, J. J. (1971b). The information available in pictures. *Leonardo*, *4*, 27–35.

Gibson, J. J. (1976). The myth of passive perception: a reply to Richards. *Philosophy and Phenomenological Research*, *37*, 234–238.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA, Houghton Mifflin.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gilchrist, A. L., & Jacobsen, A. (1983). Lightness constancy through a veiling luminance. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 936–944.

Glass, L. (1969). Moiré effect from random dots. *Nature*, *223*, 578–580.

Gordon, I. E., & Field, D. (1978). A possible explanation as to why the newly sighted commonly perform well on pseudoisochromatic colour vision tests. *Perception*, *7*, 119–122.

Gordon, I. E., & Slater, A. M. (1998). Nativism and empiricism: the history of two ideas. In A. Slater (Ed.), *Perceptual development: Visual, auditory and speech perception in infancy*. Hove, UK: Psychology Press.

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *59*, 544–549.

Gould, S. J. (1980). *The panda's thumb*. London: Penguin.

Granrud, C. E., Yonas, A., Smith, I. M., Arterberry, M. E., Glicksman, M. L., & Sorknes, A. C. (1984). Infants' sensitivity to accretion and deletion of texture as information for depth at an edge. *Child Development*, *55*, 1630–1636.

Gregory, R. L. (1963). Distortion of space as inappropriate constancy scaling. *Nature*, *199*, 678–680.

Gregory, R. L. (1974). Perceptions as hypotheses. In S. C. Brown (Ed.), *Philosophy of psychology* (Chapter 9). London: Macmillan.

Gregory, R. L. (1980a). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London*, *Series B,290*, 181–197.

Gregory, R. L. (1980b). Choosing a paradigm for perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 1). New York: Academic Press.

Gregory, R. L. (Ed.) (1987). *The Oxford companion to the mind*. Oxford: Oxford University Press.

Gregory, R. L. (1995). Black boxes of artful vision. In R. L. Gregory, J. Harris, P. Heard, & D. Rose (Eds.), *The artful eye* (Chapter 1). Oxford: Oxford University Press.

Gregory, R. L., & Wallace, J. G. (1963). Recovery from early blindness: A case study. *Experimental Psychology Society Monograph No. 2*. Cambridge: W. Heffer & Son.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.

Hammond, K. R. (Ed.) (1966). *The psychology of Egon Brunswik*. New York: Holt, Reinhart & Winston.

Hartline, H. K. (1938). The response of single optic nerve fibres of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, *121*, 400–415.

Hartline, H. K. (1940). The receptive field of the optic nerve fibres. *American Journal of Physiology*, *130*, 690–699.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Helmholtz, H. von (1924–1925). *Helmholtz's physiological optics* [Trans. from the 3rd ed. (1909–1911) by J. P. Southwell (Ed.)]. Rochester, NY: Optical Society of America.

Helson, H. (1947). Adaptation level as frame of reference for prediction of psychophysical data. *American Journal of Psychology*, *60*, 1–29.

Henle, M. (1984). Isomorphism: Setting the record straight. *Psychological Research*, *46*, 317–327.

Hering, E. (1890). Beitrag zür Lehre vom Simultankontrast. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, *1*, 18–28.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American*, *212*, 46–54.

Hess, E. H. (1975). The role of pupil size in communication. *Scientific American*, *222*, 110–119.

Hildebrandt, S., & Tromba, A. (1985). *Mathematics and optimum form*. New York: W. H. Freeman.

Hochberg, J. E. (1968). In the mind's eye. In R. N. Habet (Ed.), *Contemporary research in visual perception*. New York: Holt, Reinhart & Winston.

Hochberg, J. E. (1971). Perception. In L. A. Riggs & J. W. Kling (Eds.), *Woodworth and Schlosberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

Hochberg, J. E. (1973). Organization and the Gestalt tradition. In E. Carterette & M. Fredman (Eds.), *Handbook of perception* (Vol. 1). New York: Academic Press.

Horn, B. K. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, *3*, 277–299.

Howe, C. Q., & Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Science of the USA*, *99*, 13184–13188.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *166*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and the functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Hubel, D. H., & Wiesel, T. W. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London*, *198*, 1–59.

James, W. (1890). *Principles of psychology*. New York: Holt.

Johansson, G. (1950). *Configurations in event perception*. Stockholm: Almqvist & Wiksell.

Johansson, G. (1964). Perception of motion and changing form. *Scandinavian Journal of Psychology*, *5*, 181–208.

Johansson, G. (1977). Spatial constancy and motion in visual perception. In W. Epstein (Ed.), *Stability and constancy in visual perception*. New York: Wiley.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell System Technical Journal*, *39*, 1125–1162.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature*, *290*, 91–97.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, *20*, 525–540.

Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception*. New York: Praeger.

Katz, D. (1951). *Gestalt psychology*. London: Methuen.

Katz, S. (1987). Is Gibson a Realist? In A. Costall & A. Still (Eds.). *Cognitive psychology in question*. Brighton, UK: Harvester Press.

Kaufman, L. (1974). *Sight and mind: An introduction to visual perception*. New York: Oxford University Press.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.

Koffka, K. (1915). Reply to Benussi. Reprinted (1929) in W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Humanities Press.

Koffka, K. (1924). *The growth of the mind*. London: Routledge & Kegan Paul.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt Brace.

Kohler, I. (1955). Experiments with prolonged optical distortion. *Acta Psychologica*, *11*, 176–178.

Köhler, W. (1920). *Physical Gestalten*. Reprinted (1929) in W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Humanities Press.

Köhler, W. (1925). Reply to G. E. Müller. Reprinted (1929) in W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Humanities Press.

Köhler, W. (1940). *Dynamics in psychology*. New York: Liveright.

Köhler, W. (1947). *Gestalt psychology*. New York: Liveright.

Kolmogorov, A. N., & Uspenskii, V. A. (1987). Algorithms and randomness. *Theory of Probability and Its Applications*, *32*, 389–412.

Kubovy, M., & Pomerantz, J. T. (1981). *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*, 37–68.

Kuhn, T. S. (1970). *The nature of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Külpe, O. (1904). Versuche über Abstraktion. *Berlin International Congress of Experimental Psychology*, *56–68*.

Land, E. H. (1985). Recent advances in retinex theory. *Vision Research*, *26*, 7–21.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, *61*, 1–11.

Lashley, K. S. (1950). In search of the engram. *Symposium of the Society of Experimental Biology*, *4*, 454–482.

Lazarus, R. S., & McCleary, R. A. (1951). Autonomic discrimination without awareness: A study in subception. *Psychological Review*, *58*, 113–122.

Lee, D. N. & Reddish, P. E. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*, *293*, 293–294.

Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception and Psychophysics*, *15*, 529–532.

Lee, D. N., Reddish, P. E., & Rand, D. T. (1991). Aerial docking in hummingbirds. *Naturwissenschaften*, *78*, 526–527.

Lee, D. N., Young, D. S., & Rewt, D. (1992). How do somersaulters land on their feet? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1195–1202.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineering*, *47*, 1940–1951.

MacLeod, R. B., & Pick, H. L. (eds) (1974). *Perception: Essays in honor of James J. Gibson*. Ithaca, NY.: Cornell University Press.

MacNichol, E. (1964). Three-pigment color vision. *Scientific American*, *211*, 48–56.

Mandler, J. M., & Mandler, G. (1969). The diaspora of experimental psychologists: the Gestaltists and others. In D. Fleming & B. Bailin (Eds.), *The intellectual migration, Europe and America, 1930–1960*. Cambridge, MA: Harvard University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London*, Series B, *207*, 187–217.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*, 283–287.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, Series B, *204*, 301–328.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., & Pitts, W. H. (1960). Anatomy and physiology of vision in the frog (*Rana pipens*). *Journal of General Physiology*, *43*, 129–176.

Mayhew, J. E. W., & Frisby, J. P. (1981). Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, *17*, 349–385.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing*. Cambridge, MA: MIT Press.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review*, *56*, 244–251.

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision Research*, *42*, 1017–1033.

Meltzoff, A. N. (1995). Infants' understanding of people and things: From body imitation to folk psychology. In J. L. Bermudez, A. Marcel, & N. Eilan (Eds.), *The body and the self*. Cambridge, MA, and London: MIT Press.

Melzack, R., & Wall, P. D. (1965). Pain mechanisms: A new theory. *Science*, *150*, 971–979.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Englewood Cliffs, NJ: Prentice-Hall.

Michotte, A. (1946). *La perception de la causalité*. Louvain: Institut Superieur de Philosophie.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Møller, A. (1992). Females prefer large and symmetrical ornaments. *Nature*, *357*, 238–240.

Møller, A. (1994). Sexual selection in the barn swallow (*Hirundo rustica*). VI. Patterns of fluctuating asymmetry and selection against asymmetry. *Evolution*, *48*, 658–670.

Møller, A. (1995). Bumblebee preference for symmetrical flowers. *Proceedings of the National Academy of Science of the USA*, *92*, 2288–2292.

Møller, A., Soler, M., & Thornhill, R. (1995). Breast asymmetry, sexual selection, and human reproductive success. *Ethology and Sociobiology*, *16*, 207–209.

Morgan, M. J. (1977). *Molyneux's question: Vision, touch and the philosophy of perception*. Cambridge: Cambridge University Press.

Morgan, M. J. (1984). Computational theories of vision (review of Marr). *Quarterly Journal of Experimental Psychology*, *36A*, 157–165.

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

Nelson, B. (2000). *The Atlantic gannet* (2nd ed.). Great Yarmouth: Fenix Books Ltd.

Noble, W. G. (1981). Gibsonian theory and the pragmatist perspective. *Journal for the Theory of Social Behaviour*, *11*, 65–85.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1992). *Turn signals are the facial expressions of automobiles*. Reading, MA: Addison-Wesley.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wesley.

Norman, D. A., & Draper, S. W. (Eds.) (1986). *User-centred system design*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*, 73–144.

O'Brien, G. O., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioural and Brain Sciences*, *22*, 127–196.

Orchard, G. A., & Phillips, W. A. (1991). *Neural computation: A beginner's guide*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Pantle, A., & Sekuler, R. W. (1968). Contrast response of human visual mechanisms to orientation and detection of velocity. *Vision Research*, *9*, 397–406.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex: A clinical study of localization of function*. New York: Macmillan.

Petermann, B. (1932). *The Gestalt theory and the problem of configuration*. London: Kegan Paul, Trench & Trubner.

Petrinovich, L. (1979). Probabilistic functionalism: A conception of research method. *American Psychologist*, *34*, 373–390.

Petter, G. (1956). Nuove ricerche sperimentali sulla totalizzazione percettiva. *Rivista di Psicologia*, *50*, 213–227.

Piaget, J. (1954). *The construction of reality in the child*. London: Routledge & Kegan Paul.

Pittenger, J. B., & Shaw, R. E. (1975). Ageing faces as viscal-elastic events. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 374–382.

Pomerleau, D. A. (1989). ALVINN: An autonomous land vehicle in a neural network. In D. S. Touretzky (Ed.), *Advances in neural information processing systems*. New York: Morgan Kaufmann.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Popper, K. R. (1960). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge & Kegan Paul.

Porter, P. B. (1954). Find the hidden man. *American Journal of Psychology*, *67*, 550–551.

Purves, D., & Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinaur Associates Inc.

Quinn, P. C., Burke, S., & Rush, A. (1993). Part–whole perception in early infancy: Evidence for perceptual grouping produced by lightness similarity. *Infant Behavior and Development*, *16*, 19–42.

Ramachandran, V. S. (1990). Interactions between motion, depth, color, and form: the utilitarian theory of perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge: Cambridge University Press.

Reed, E. (1982). Descartes' corporeal ideas hypothesis and the origin of scientific psychology. *Review of Metaphysics*, *35*, 731–752.

Reed, E. S. (1987). James Gibson's ecological approach to cognition. In A. Costall and A. Still (Eds.), *Cognitive psychology in question*. Brighton, UK: Harvester Press.

Restle, F. (1979). Coding theory of the perception of motion configuration. *Psychological Review*, *86*, 1–24.

Rock, I. (1995). *Perception* (2nd ed.). New York: Scientific American Books.

Rogers, S., & Costall, A. (1983). On the horizon: Picture perception and Gibson's concept of information. *Leonardo*, *16*, 180–182.

Rosenblatt, F. (1959). Two theorems of statistical separability. In *Mechanisation of thought processes*. Proceedings of a symposium held at the National Physical Laboratory, 1 November. London: HMSO.

Rosenzweig, M. R., Leiman, A. L., & Breedlove, S. M. (1999). *Biological psychology: An introduction to behavioral, cognitive, and clinical neuroscience* (2nd ed.). Sunderland, MA: Sinaur Associates Inc.

Rubin, E. (1915). *Synsoplevede figurer*. Copenhagen: Glyndendalska.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.

Rushton, W. A. (1964). Color blindness and cone pigments. *American Journal of Optometry and Archives of the American Academy of Optometry*, *41*, 265–282.

Russell, J. (1984). *Explaining mental life*. London: Macmillan.

Sanford, R. H. (1936). The effects of abstinence from food upon imaginal processes: A preliminary experiment. *Journal of Psychology*, *2*, 129–136.

Sedgwick, H. A. (1980). The geometry of spatial layout in pictorial representation. In M. Hagen (Ed.), *The perception of pictures* (Vol. 1, Chapter 2). New York: Academic Press.

Sekuler, R., & Blake, R. (1985). *Perception*. New York: McGraw-Hill.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *The mechanisms of thought processes*. London: HMSO.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American*, *203*, 60–68.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*, 2547–2554.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, *27*, 379–425.

Shaw, R., McIntyre, M., & Mace, W. (1974). The role of symmetry in event perception. In R. B. MacLeod & H. L. Pick (Eds.), *Perception: Essays in honor of James J. Gibson*. Ithaca, NY: Cornell University Press.

Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, *91*, 417–447.

Shepard, R. N. (1990). *Mind sights*. New York: W. H. Freeman.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The advance in operations research. *Operations Research*, *6*, 6.

Skottun, B. C., De Valois, R. L., Grosof, D. H., Movshon, A., Albrech, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Research*, *31*(7/8), 1079–1086.

Slater, A. (2001). Visual perception. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (Chapter 1). Oxford: Blackwell.

Slater, A. (2002). The development of the senses. In A. Slater & M. Lewis (Eds.), *Introduction to infant development* (Chapter 5). Oxford: Oxford University Press.

Slater, A., & Morison, V. (1985). Shape constancy and slant perception at birth. *Perception*, *14*, 337–344.

Slater, A., Bremner, J. G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. (2000). Newborn infants' preference for attractive faces: The role of internal and external facial features. *Infancy*, *1*, 265–274.

Slater, A., Field, T., & Hernandez-Reif, M. (2002). The development of the senses. In A. Slater & M. Lewis (Eds.), *Introduction to infant development*. Oxford: Oxford University Press.

Slater, A., Mattock, A., & Browne, E. (1990). Size constancy at birth: Newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology*, *49*, 314–322.

Slater, A., Rose, D., & Morison, V. (1984). Newborn infants' perception of similarities and differences between two- and three-dimensional stimuli. *British Journal of Developmental Psychology*, *2*, 287–294.

Solomonoff, R. J. (1964). A formal theory of inductive inference, Parts 1 and 2. *Information and Control*, *7*, 1–22, 224–254.

Sommer, R. (1959). The new look on the witness stand. *Canadian Psychologist*, *8*, 94–99.

Spence, K. W. (1956). *Behavior theory and conditioning*. New Haven, CT: Yale University Press.

Sperry, R. W. (1951). Mechanisms of neural maturation. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.

Sperry, R. W., & Miner, W. (1955). Pattern perception following insertion of mica plates into visual cortex. *Journal of Comparative and Physiological Psychology*, *48*, 463–469.

Sperry, R. W., Miner, W., & Meyers, R. E. (1955). Visual pattern perception following subpial string and tantalum wire implantations in the visual cortex. *Journal of Comparative and Physiological Psychology*, *48*, 50–58.

Staniland, A. (1966). *Patterns of redundancy*. Cambridge: Cambridge University Press.

Stevens, K. A. (1978). Computation of locally parallel structure. *Biological Cybernetics*, *29*, 19–28.

Sutherland, N. S. (1957). Visual discrimination of shape by octopus. *British Journal of Psychology*, *48*, 55–70.

Svaetichin, G. (1956). Spectral response curves from single cones. *Acta Physiologica Scandinavica Supplementum*, *134*, 17–46.

Tinbergen, N. (1951). *The study of instinct*. London: Oxford University Press.

Titchener, E. B. (1901). *Experimental psychology: A manual of laboratory practice*. New York: Macmillan.

Uhr, L. (1963). Pattern recognition computers as models for form perception. *Psychological Bulletin*, *60*, 40–73.

Ullman, S. (1980). Against direct perception. *The Behavioral and Brain Sciences*, *3* (whole issue).

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In J. D. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, MA: MIT Press.

Valvo, A. (1971). *Sight restoration after long-term blindness: The problems and behavior patterns of visual rehabilitation*. New York: American Foundation for the Blind.

van der Helm, P. A. (2000). Simplicity vs. likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*(5), 770–800.

von Hofsten, C. (1983). Foundations for perceptual development. In L. P. Lipsitt & C. K. Rovee-Collier (Eds.). *Advances in infancy research* (Vol. 2). Norwood, NJ: Ablex.

von Senden, M. (1960). *Space and sight*. London: Methuen.

Walls, G. L. (1942). *The vertebrate eye and its adaptive radiation*. Birmingham, MI: Cranbrook Institute of Science.

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 683–703.

Watt, R. J. (1988). *Visual processing: Computational, psychophysical and cognitive research*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, *25*, 1661–1674.

Wertheimer, M. (1912). Experimental studies on the seeing of motion. Reprinted (1961) in T. Shipley, *Classics in psychology*. New York: Philosophical Library.

Wertheimer, M. (1923). Untersuchungen zur Lehrer von der Gestalt. Reprinted (1961) in T. Shipley, *Classics in psychology*. New York: Philosophical Library.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwood, NJ: Ablex.

Zaal, F. T. J. M., & Michaels, C. F. (2003). The information for catching fly balls: Judging and intercepting virtual balls in a CAVE. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 537–555.

Zeki, S. (1993). *A vision of the brain*. Oxford: Blackwell.

Author index

Page numbers in *italic* indicate figures

Subject index

Page numbers in *italic* indicate figures