Approaches to human cognition

INTRODUCTION

We are now well into the third millennium and there is ever-increasing interest in unravelling the mysteries of the human brain and mind. This interest is reflected in the substantial upsurge of scientific research within cognitive psychology and cognitive neuroscience. In addition, the cognitive approach has become increasingly influential within clinical psychology. In that area, it is recognised that cognitive processes (especially cognitive biases) play a major role in the development (and successful treatment) of mental disorders (see Chapter 15).

In similar fashion, social psychologists increasingly focus on **social cognition**. This focuses on the role of cognitive processes in influencing individuals' behaviour in social situations. For example, suppose other people respond with laughter when you tell them a joke. This laughter is often ambiguous – they may be laughing with you or at you (Walsh et al., 2015). Your subsequent behaviour is likely to be influenced by your cognitive interpretation of their laughter.

What *is* cognitive psychology? It is concerned with the internal processes involved in making sense of the environment and deciding on appropriate action. These processes include attention, perception, learning, memory, language, problem solving, reasoning and thinking. We can define **cognitive psychology** as aiming to understand human cognition by observing the behaviour of people performing various cognitive tasks. However, the term "cognitive psychology" can also be used more broadly to include brain activity and structure as relevant information for understanding human cognition. It is in this broader sense that it is used in the title of this book.

Here is a simple example of cognitive psychology in action. Frederick (2005) developed a test (the Cognitive Reflection Test) that included the following item:

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ___ cents

KEY TERMS

Social cognition
An approach within social psychology in which the emphasis is on the cognitive processing of information about other people and social situations.

Cognitive psychology
An approach that aims to understand human cognition by the study of behaviour; a broader definition also includes the study of brain activity and structure.

> **KEY TERM** 🔑
>
> **Cognitive neuroscience**
> An approach that aims to understand human cognition by combining information from behaviour and the brain.

What do *you* think is the correct answer? Brañas-Garza et al. (2015) found in a review of findings from 41,004 individuals that 68% produced the wrong answer (typically 10 cents) and only 32% gave the right answer (5 cents). Even providing financial incentives to produce the correct answer failed to improve performance.

The above findings suggest most people will rapidly produce an incorrect answer (i.e., 10 cents) that is easily accessible and are unwilling to devote extra time to checking that they have the right answer. However, Gangemi et al. (2015) found many individuals producing the wrong answer had a feeling of error suggesting they experienced cognitive uneasiness about their answer. In sum, the intriguing findings on the Cognitive Reflection Test indicate that we can fail to think effectively even on relatively simple problems. Subsequent research has clarified the reasons for these deficiencies in our thinking (see Chapter 12).

The aims of cognitive neuroscientists overlap with those of cognitive psychologists. However, there is one major difference between cognitive neuroscience and cognitive psychology in the narrow sense. Cognitive neuroscientists argue convincingly we need to study the *brain* as well as behaviour while people engage in cognitive tasks. After all, the internal processes involved in human cognition occur in the brain. Cognitive neuroscience uses information about behaviour and the brain to understand human cognition. Thus, the distinction between cognitive neuroscience and cognitive psychology in the broader sense is blurred.

Cognitive neuroscientists explore human cognition in several ways. First, there are brain-imaging techniques of which functional magnetic resonance imaging (fMRI) is probably the best-known. Second, there are electrophysiological techniques involving the recording of electrical signals generated by the brain. Third, many cognitive neuroscientists study the effects of brain damage on cognition. It is assumed the patterns of cognitive impairment shown by brain-damaged patients can inform us about normal cognitive functioning and the brain areas responsible for various cognitive processes.

The huge increase in scientific interest in the workings of the brain is mirrored in the popular media – numerous books, films and television programmes communicate the more accessible and dramatic aspects of cognitive neuroscience. Increasingly, media coverage includes coloured pictures of the brain indicating the areas most activated when people perform various tasks.

Four main approaches

We can identify four main approaches to human cognition (see Table 1.1). Note, however, there has been a substantial increase in research combining two (or even more) of these approaches. We will shortly discuss each approach in turn and you will probably find it useful to refer back to this chapter when reading the rest of the book. Hopefully, you will find Table 1.3 (towards the end of this chapter) especially useful because it summarises the strengths and limitation of all four approaches.

TABLE 1.1 APPROACHES TO HUMAN COGNITION

1. *Cognitive psychology*: this approach involves using behavioural evidence to enhance our understanding of human cognition. Since behavioural data are also of great importance within cognitive neuroscience and cognitive neuropsychology, cognitive psychology's influence is enormous.

2. *Cognitive neuropsychology*: this approach involves studying brain-damaged patients to understand normal human cognition. It was originally closely linked to cognitive psychology but has recently also become linked to cognitive neuroscience.

3. *Cognitive neuroscience*: this approach involves using evidence from behaviour and the brain to understand human cognition.

4. *Computational cognitive science*: this approach involves developing computational models to further our understanding of human cognition; such models increasingly incorporate knowledge of behaviour and the brain. A computational model takes the form of an **algorithm**, which consists of a precise and detailed specification of the steps involved in performing a task. Computational models are designed to simulate or imitate human processing on a given task.

COGNITIVE PSYCHOLOGY

We can obtain some perspective on the contribution of cognitive psychology by considering what preceded it. Behaviourism was the dominant approach to psychology throughout the first half of the twentieth century. The American psychologist John Watson (1878–1958) is often regarded as the founder of behaviourism. He argued that psychologists should focus on stimuli (aspects of the immediate situation) and responses (behaviour produced by the participants in an experiment). This approach appears "scientific" because it focuses on stimuli and responses, both of which are observable.

Behaviourists argued that internal mental processes (e.g., attention) cannot be verified by reference to observable behaviour and so should be ignored. According to Watson (1913, p. 165), behaviourism should "never use the terms consciousness, mental states, mind, content, introspectively verifiable and the like". In stark contrast, as we have already seen, cognitive psychologists argue it is of crucial importance to study such internal mental processes. Hopefully, you will be convinced that cognitive psychologists are correct when you read how the concepts of attention (Chapter 5) and consciousness (Chapter 16) have been used fruitfully to enhance our understanding of human cognition.

It is often claimed that behaviourism was overthrown by the "cognitive revolution". However, the reality was less dramatic (Hobbs & Burman, 2009). For example, Tolman (1948) was a behaviourist but he did not believe internal processes should be ignored. He carried out studies in which rats learned to run through a maze to a goal box containing food. When Tolman blocked off the path the rats had learned to use, they rapidly learned to follow other paths leading in the right general direction. Tolman concluded the rats had acquired an internal *cognitive map* indicating the maze's approximate layout.

It is almost as pointless to ask "When did cognitive psychology start?", as to enquire "How long is a piece of string?". However, 1956 was crucially

KEY TERMS 🔑

Bottom-up processing
Processing directly influenced by environmental stimuli; see **top-down processing**.

Serial processing
Processing in which one process is completed before the next one starts; see **parallel processing**.

Top-down processing
Stimulus processing that is influenced by factors such as the individual's past experience and expectations.

important. At a meeting at the Massachusetts Institute of Technology, Noam Chomsky presented his theory of language, George Miller discussed the magic number seven in short-term memory (Miller, 1956) and Alan Newell and Herbert Simon discussed the General Problem Solver (see Gobet and Lane, 2015). In addition, there was the first systematic attempt to study concept formation from the cognitive perspective (Bruner et al., 1956). The history of cognitive psychology from the perspective of its classic studies is discussed in Eysenck and Groome (2015a).

Several decades ago, most cognitive psychologists subscribed to the information-processing approach based loosely on an analogy between the mind and the computer (see Figure 1.1). A stimulus (e.g., a problem or task) is presented, which causes various internal processes to occur, leading eventually to the desired response or answer. Processing directly affected by the stimulus input is often described as **bottom-up processing**. It was typically assumed only one process occurs at a time: this is **serial processing**, meaning the current process is completed before the onset of the next one.
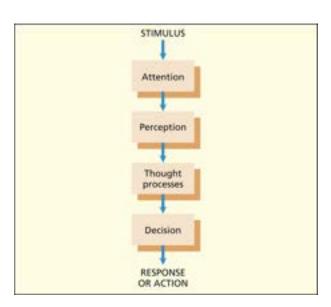

Figure 1.1
An early version of the information processing approach.

Figure 1.2
Diagram to demonstrate top–down processing.

The above approach is drastically over-simplified. Task processing typically also involves **top-down processing**, which is processing influenced by the individual's expectations and knowledge rather than simply by the stimulus itself. Read what it says in the triangle (Figure 1.2). Unless you know the trick, you probably read it as "Paris in the spring". If so, look again: the word "the" is repeated. Your expectation it was a well-known phrase (i.e., top-down processing) dominated the information available from the stimulus (i.e., bottom-up processing).

The traditional approach was also over-simplified in assuming processing is typically serial. In fact, more than one process typically occurs at the same time – this is **parallel processing**. We are much more likely to use parallel processing when performing a highly practised task than a new one (see Chapter 5). For example, someone taking their first driving lesson finds it very hard to control the car's speed, steer accurately and pay attention to other road users at the same time. In contrast, an experienced driver finds it easy.

There is also **cascade processing**: a form of parallel processing involving an *overlap* of different processing stages when someone performs a task. More specifically, later stages of processing are initiated before one or more earlier stages have finished. For example, suppose you are trying to work out the meaning of a visually presented word.

The most thorough approach would involve identifying all the letters in the word followed by matching the resultant letter string against words you have stored in long-term memory. In fact, people often engage in cascade processing – they form hypotheses as to the word that has been presented *before* identifying all the letters (McClelland, 1979).

An important issue for cognitive psychologists is the task-impurity problem – most cognitive tasks require several processes thus making it hard to interpret the findings. One approach to this problem is to consider various tasks all requiring the same process. For example, Miyake et al. (2000) used three tasks requiring deliberate inhibition of a dominant response:

(1) The Stroop task: name the colour in which colour words are presented (e.g., RED printed in green) and avoid saying the colour word (which has to be inhibited). You can see for yourself how hard this task is by naming the colours of the words shown in Figure 1.3.
(2) The anti-cascade task: inhibit the natural tendency to look at a visual cue and instead look in the opposite direction. People typically take longer to perform this task than the control task of simply looking at the visual cue.
(3) The stop-signal task: respond rapidly to indicate whether each of a series of words is an animal or non-animal; on key trials, there was a computer-emitted tone indicating that the response should be inhibited.

Miyake et al. (2000) found all three tasks involved similar processes. They used complex statistical techniques (latent variable analysis) to extract what
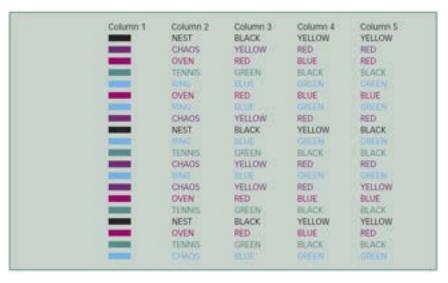

Figure 1.3
Test yourself by naming the colours in each column. You should name the colours rapidly in the first three columns because there is no colour-word conflict. In contrast, colour naming should be slower (and more prone to error) when naming colours in the fourth and fifth columns.

KEY TERMS 🔑

Parallel processing
Processing in which two or more cognitive processes occur at the same time.

Cascade processing
Later processing stages start before earlier processing stages have been completed when performing a task.

was common across the three tasks. This was assumed to represent a relatively pure measure of the inhibitory process. Throughout this book, we will discuss many ingenious strategies used by cognitive psychologists to identify the processes used in numerous tasks.

Strengths

Cognitive psychology was for many years the engine room of progress in understanding human cognition and the other three approaches listed in Table 1.1 have benefitted from it. For example, cognitive neuropsychology became important 25 years after cognitive psychology. It was only when cognitive psychologists had developed reasonable accounts of healthy human cognition that the performance of brain-damaged patients could be understood fully. Before that, it was hard to decide which patterns of cognitive impairment were theoretically important.

In a similar fashion, the computational modelling activities of computational cognitive scientists are typically heavily influenced by pre-computational psychological theories. Finally, the great majority of theories driving research in cognitive neuroscience originated within cognitive psychology.

Cognitive psychology has not only had a massive influence on theorising across all four major approaches to human cognition. It has also had a predominant influence on the development of cognitive tasks and on task analysis (how a task is accomplished).

Limitations

In spite of cognitive psychology's enormous contributions, it has several limitations. First, our behaviour in the laboratory may differ from our behaviour in everyday life. Thus, laboratory research sometimes lacks ecological validity – the extent to which laboratory findings are applicable to everyday life. For example, our everyday behaviour is often designed to change a situation or to influence others' behaviour. In contrast, the sequence of events in most laboratory research is based on the experimenter's predetermined plan and is uninfluenced by participants' behaviour. Wachtel (1973) used the term implacable experimenter to describe this state of affairs.

We must not exaggerate problems associated with lack of ecological validity. As we will see in this book, there has been a dramatic increase in applied cognitive psychology in which the emphasis is on investigating topics of general importance. Such research often has good ecological validity. Note that it is far better to carry out well-controlled experiments under laboratory conditions than poorly controlled experiments under naturalistic conditions. It is precisely because it is considerably easier for researchers to exercise experimental control in the laboratory that so much research is laboratory-based.

Second, theories in cognitive psychology are often expressed only in verbal terms (although this is becoming less common). Such theories are vague, making it hard to know precisely what predictions follow from them and thus to falsify them. These limitations can largely be overcome by

computational cognitive scientists developing cognitive models specifying precisely any given theory's assumptions.

Third, difficulties in falsifying theories have led to a proliferation of different theories on any given topic. For example, there are at least 12 different theories of working memory (see Chapter 6). Another reason for the proliferation of rather similar theories is the "toothbrush problem" (Mischel, 2008): no self-respecting cognitive psychologist wants to use anyone else's theory.

Fourth, the findings obtained using any given task or paradigm are sometimes *specific* to that paradigm and do not generalise to other (apparently similar) tasks. This is paradigm specificity. It means some findings are narrow in scope and applicability (Meiser, 2011). This problem can be minimised by developing theories accounting for performance across several tasks or paradigms. For example, Anderson et al. (2004; discussed later in this chapter) developed a comprehensive theoretical architecture or framework known as the Adaptive Control of Thought-Rational (ACT-R) model.

Fifth, cognitive psychologists typically obtain measures of performance speed and accuracy. These measures are very useful but provide only *indirect* evidence about internal cognitive processes. Most tasks are "impure" in that they involve several processes, and it is hard to identify the number and nature of processes involved on the basis of speed and accuracy measures.

COGNITIVE NEUROPSYCHOLOGY

Cognitive neuropsychology focuses on the patterns of cognitive performance (intact and impaired) of brain-damaged patients having a lesion (structural damage to the brain caused by injury or disease). According to cognitive neuropsychologists, studying brain-damaged patients can tell us much about cognition in healthy individuals.

The above idea does not sound very promising, does it? In fact, however, cognitive neuropsychology has contributed substantially to our understanding of healthy human cognition. For example, in the 1960s, most memory researchers thought the storage of information in long-term memory depended on previous processing in short-term memory (see Chapter 6). However, Shallice and Warrington (1970) reported the case of a brain-damaged man, KF. His short-term memory was severely impaired but his long-term memory was intact. These findings played an important role in changing theories of healthy human memory.

Since cognitive neuropsychologists study brain-damaged patients, we might imagine they would be interested in the workings of the brain. In fact, many cognitive neuropsychologists pay little attention to the brain itself. According to Coltheart (2015, p. 198), for example, "Even though cognitive neuropsychologists typically study people with brain damage, . . . cognitive neuropsychology is not about the brain: it is about information-processing models of cognition."

An increasing number of cognitive neuropsychologists disagree with Coltheart. They believe we should consider the brain, using techniques such as magnetic resonance imaging to identify the brain areas damaged in any given patient. They are also increasingly willing to study the impact of brain damage on brain processes using various neuroimaging techniques.

<div style="float:right;border:1px solid #888;padding:8px;">

KEY TERMS 🔑

Paradigm specificity
The findings with a given experimental task or paradigm are not replicated even when apparently very similar tasks or paradigms are used.

lesion
Damage within the brain resulting from injury or disease; it typically affects a restricted area.

</div>

Max Coltheart.
Courtesy of Max Coltheart.

Theoretical assumptions

Coltheart (2001) provided a very clear account of the major assumptions of cognitive neuropsychology. Here we will discuss these assumptions and briefly consider relevant evidence.

One key assumption is modularity, meaning the cognitive system consists of numerous modules or processors operating fairly independently or separately of each other. It is assumed these modules exhibit domain specificity (they respond to only one given class of stimuli). For example, there may be a face-recognition module that responds only when a face is presented.

Modular systems typically involve serial processing with processing within one module being completed before processing starts in the next module. As a result, there is very limited *interaction* among modules. There is some support for modularity from the evolutionary approach. Species with larger brains generally have more specialised brain regions that could be involved in modular processing. However, the notion that human cognition is heavily modular is hard to reconcile with neuroimaging evidence. The human brain possesses a moderately high level of connectivity (Bullmore & Sporns, 2012; see p. 14), suggesting there is more parallel processing than assumed by most cognitive neuropsychologists.

The second major assumption is that of *anatomical modularity*. According to this assumption, each module is located in a specific brain area. Why is this assumption important? Cognitive neuropsychologists are most likely to make progress when studying brain patients with brain damage limited to a *single* module. Such patients may not exist if there is no anatomical modularity. Suppose all modules were distributed across large brain areas. If so, the great majority of brain-damaged patients would suffer damage to most modules, making it impossible to work out the number and nature of their modules.

There is evidence of anatomical modularity in the visual processing system (see Chapter 2). However, there is less support for anatomical modularity with most complex tasks. For example, consider the findings of Yarkoni et al. (2011). Across over 3,000 neuroimaging studies, some brain areas (e.g., dorsolateral prefrontal cortex; anterior cingulate cortex) were activated in 20% of them despite the great diversity of tasks involved.

The third major assumption (the "universality assumption") is that "Individuals . . . share a similar or an equivalent organisation of their cognitive functions, and presumably have the same underlying brain anatomy"

> **KEY TERM** 🔑
>
> **Modularity**
> The assumption that the cognitive system consists of many fairly independent or separate modules or processors, each specialised for a given type of processing.

(de Schotten and Shallice, 2017, p. 172). If this assumption (also common within cognitive neuroscience) is false, we could not readily use the findings from individual patients to draw conclusions about the organisation of other people's cognitive systems or functional architecture.

There is accumulating evidence against the universality assumption. Tzourio-Mazoyer et al. (2004) discovered substantial differences between individuals in the location of brain networks involved in speech and language. Finn et al. (2015) found clear-cut differences between individuals in functional connectivity across the brain, concluding that "An individual's functional brain connectivity profile is both unique and reliable, similarly to a fingerprint" (p. 1669).

Duffau (2017) reviewed interesting research conducted on patients during surgery for epilepsy or a tumour. Direct electrical stimulation, which causes "a genuine virtual transient lesion" (p. 305) is applied invasively to the cortex. The patient is awakened and given various cognitive tasks while receiving stimulation. Impaired performance when direct electrical stimulation is applied to a given area indicates that area is involved in the cognitive functions assessed by the current task.

Findings obtained using direct electrical stimulation and other techniques (e.g., fMRI) led Duffau (2017) to propose a two-level model. At the *cortical* level, there is high variability across individuals in structure and function of any given brain areas. At the *subcortical* level (e.g., in premotor cortex), in contrast, there is very little variability across individuals. The findings at the cortical level seem inconsistent with the universality assumption.

The fourth assumption is *subtractivity*. The basic idea is that brain damage impairs one or more processing modules but does not change or add anything. The fifth assumption (related to subtractivity) is *transparency* (Shallice, 2015). According to the transparency assumption, the performance of a brain-damaged patient reflects the operation of a theory designed to explain the performance of healthy individuals minus the impact of their lesion.

Why are the subtractivity and transparency assumptions important? Suppose they are incorrect and brain-damaged patients develop new modules to compensate for their cognitive impairments. That would greatly complicate the task of learning about the intact cognitive system by studying brain-damaged patients. Consider pure alexia, a condition in which brain-damaged patients have severe reading problems but otherwise intact language abilities. These patients generally have a direct relationship between word length and reading speed due to letter-by-letter processing (Bormann et al., 2015). This indicates the use of a compensatory strategy differing markedly from the reading processes used by healthy adults.

Research in cognitive neuropsychology

How do cognitive neuropsychologists set about understanding the cognitive system? Of major importance is the search for dissociations, which occur when a patient has normal performance on one task (task X) but is impaired on a second one (task Y). For example, amnesic patients perform almost normally on short-term memory tasks but are greatly impaired on many

KEY TERMS 🔑

Double dissociation
The finding that some brain-damaged individuals have intact performance on one task but poor performance on another task whereas other individuals exhibit the opposite pattern.

Association
The finding that certain symptoms or performance impairments are consistently found together in numerous brain-damaged patients.

Syndrome
The notion that symptoms that often co-occur have a common origin.

Case-series study
A study in which several patients with similar cognitive impairments are tested; this allows consideration of individual data and of variation across individuals.

long-term memory tasks (see Chapter 6). It is tempting (but dangerous!) to conclude that the two tasks involve different processing modules and that the module(s) needed on long-term memory tasks have been damaged by brain injury.

Why must we avoid drawing sweeping conclusions from dissociations? Patients may perform well on one task but poorly on a second one simply because the second task is more complex. Thus, dissociations may reflect differences in task complexity rather than the use of different modules.

One apparent solution to the above problem is to find double dissociations. A **double dissociation** between two tasks (X and Y) is obtained when one patient performs normally on task X and is impaired on task Y but another patient shows the opposite pattern. We cannot explain double dissociations by arguing that one task is harder. For example, consider the double dissociation that amnesic patients have impaired long-term memory but intact short-term memory whereas other patients (e.g., KF discussed above) have the opposite pattern. This double dissociation strongly suggests there is an important distinction between short-term and long-term memory and that they involve different brain regions.

The approach based on double dissociations has various limitations. First, it is generally based on the assumption that separate modules exist (which may be misguided). Second, double dissociations can often be explained in various ways and so provide only *indirect* evidence for separate modules underlying each task (Davies, 2010). For example, a double dissociation between tasks X and Y implies the cognitive system used on X is not identical to the one used on Y. Strictly speaking, the most we can generally conclude is that "Each of the two systems has at least one sub-system that the other doesn't have" (Bergeron, 2016, p. 818). Third, it is hard to decide which of the very numerous double dissociations that have been discovered are theoretically important.

Finally, we consider associations. An **association** occurs when a patient is impaired on tasks X and Y. Associations are sometimes taken as evidence for a **syndrome** (sets of symptoms or impairments often found together). However, there is a serious flaw in the syndrome-based approach. An association may be found between tasks X and Y because the mechanisms on which they depend are adjacent anatomically in the brain rather than because they depend on the same underlying mechanism. Thus, the interpretation of associations is fraught with difficulty.

Single case studies vs case series

For many years after the rise of cognitive neuropsychology in the 1970s, most cognitive neuropsychologists made extensive use of single-case studies. There were two main reasons. First, researchers can often gain access to only one patient having a given pattern of cognitive impairment. Second, it was often assumed every patient has a somewhat different pattern of cognitive impairment and so is unique. As a result, it would be misleading and uninformative to average the performance of several patients.

In recent years, there has been a move towards the **case-series study**. Several patients with similar cognitive impairments are tested. After that,

the data of individual patients are compared and variation across patients assessed.

The case-series approach is generally preferable to the single-case approach for various reasons (Lambon Ralph et al., 2011; Bartolomeo et al., 2017). First, it provides much richer data. With a case series, we can assess the *extent of variation* between patients rather than simply being concerned about the impairment (as in the single-case approach). Second, with a case series, we can identify (and then de-emphasise) the findings from patients who are "outliers". With the single-case approach, in contrast, we do not know whether the one and only patient is representative of patients with that condition or is an outlier.

Strengths

Cognitive neuropsychology has several strengths. First, it has the advantage that it allows us to draw *causal* inferences about the relationship between brain areas and cognitive processes and behaviour. In other words, we can conclude (with moderate but not total confidence) that a given brain area is crucially involved in performing certain cognitive tasks (Genon et al., 2018).

Second, as Shallice (2015, pp. 387–388) pointed out, "A key intellectual strength of neuropsychology . . . is its ability to provide evidence falsifying plausible cognitive theories." Consider patients reading visually presented words and non-words aloud. We might imagine patients with damage to language areas would have problems in reading *all* words and non-words. However, some patients perform reasonably well when reading regular words (with predictable pronunciations) or non-words, but poorly when reading irregular words (words with unpredictable pronunciations). Other patients can read regular words but have problems with unfamiliar words and non-words. These fascinating patterns of impairment have transformed theories of reading (Coltheart, 2015; see Chapter 9).

Third, cognitive neuropsychology "produces large-magnitude phenomena which can be initially theoretically highly counterintuitive" (Shallice, 2015, p. 405). For example, amnesic patients typically have severely impaired long-term memory for personal events and experiences but an essentially intact ability to acquire and retain motor skills (Chapter 7). These strong effects played a major role in memory researchers abandoning the notion of a single long-term memory system and replacing it with more complex theories.

Fourth, in recent years, cognitive neuropsychology has increasingly been combined fruitfully with cognitive neuroscience. For example, cognitive neuroscience has revealed that a given brain injury or lesion often has widespread effects within the brain. This phenomenon is known as diaschisis: "the distant neurophysiological changes directly caused by a focal injury . . . these changes should correlate with behaviour" (Carrera & Tononi, 2014, p. 2410). Discovering the true extent of the brain areas adversely affected by a lesion facilitates the task of relating brain functioning to cognitive processing and task performance.

> **KEY TERM** 🔑
>
> **Diaschisis**
> The disruption to distant brain areas caused by a localised brain injury or lesion.

Limitations

What are the *limitations* of the cognitive neuropsychological approach? First, the crucial assumption that the cognitive system is fundamentally modular is reasonable but too strong. There is less evidence for modularity among higher-level cognitive processes (e.g., consciousness; focused attention) than among lower-level processes (e.g., colour processing; motion processing). If the modularity assumption is incorrect, this has implications for the whole enterprise of cognitive neuropsychology (Patterson & Plaut, 2009).

Second, other theoretical assumptions also seem too extreme. For example, evidence discussed earlier casts considerable doubts on the assumption of anatomical modularity and the universality assumption.

Third, the common assumption that the task performance of patients provides relatively direct evidence concerning the impact of brain damage on previously intact cognitive systems is problematic. Brain-damaged patients often make use of *compensatory strategies* to reduce or eliminate the negative effects of brain damage on cognitive performance. We saw an example of such compensatory strategies earlier – patients with pure alexia manage to read words by using a letter-by-letter strategy rarely used by healthy individuals.

Hartwigsen (2018) proposed a model to predict when compensatory processes will and will not be successful. According to this model, *general* processes (e.g., attention; cognitive control; error monitoring) can be used to compensate for the disruption of *specific* processes (e.g., phonological processing) by brain injury. However, specific processes cannot be used to compensate for the disruption of general processes. Hartwigsen discussed evidence supporting his model.

Fourth, lesions can alter the organisation of the brain in several ways. Dramatic evidence for brain plasticity is discussed in Chapter 16. Patients whose entire left brain hemisphere was removed at an early age (known as hemispherectomy) often develop good language skills even though language is typically centred in the left hemisphere (Blackmon, 2016).

There is the additional problem that a brain lesion can lead to changes in the functional connectivity between the area of the lesion and distant, intact brain areas (Bartolomeo et al., 2017). Thus, impaired cognitive performance following brain damage may reflect widespread reduced brain connectivity as well as direct damage to a specific brain area. This complicates the task of interpreting the findings obtained from brain-damaged patients.

COGNITIVE NEUROSCIENCE: THE BRAIN IN ACTION

Cognitive neuroscience involves the intensive study of brain activity as well as behaviour. Alas, the brain is extremely complicated (to put it mildly!). It consists of 100 billion neurons connected in very complex ways. We must consider how the brain is organised and how the different areas are described to understand research involving functional neuroimaging. Below we discuss various ways of describing specific brain areas.

KEY TERMS 🔑

Sulcus
A groove or furrow in the surface of the brain.

Gyrus
Prominent elevated area or ridge on the brain's surface; "gyri" is the plural.

Dorsal
Towards the top.

Ventral
Towards the bottom.

Rostral
Towards the front of the brain.

Posterior
Towards the back of the brain.

Lateral
Situated at the side of the brain.
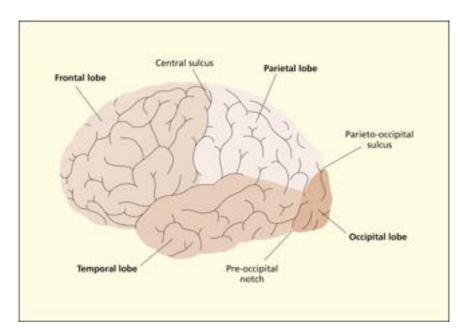
Medial
Situated in the middle of the brain.

Interactive feature:
Primal Pictures'
3D atlas of the brain

First, the cerebral cortex is divided into four main divisions or lobes (see Figure 1.4). There are four lobes in each brain hemisphere: frontal; parietal; temporal; and occipital. The frontal lobes are divided from the parietal lobes by the central **sulcus** (sulcus means furrow or groove), and the lateral fissure separates the temporal lobes from the parietal and frontal lobes. In addition, the parieto-occipital sulcus and pre-occipital notch divide the occipital lobes from the parietal and temporal lobes. The main gyri (or ridges; **gyrus** is the singular) within the cerebral cortex are shown in Figure 1.4.

Researchers use various terms to describe accurately the brain area(s) activated during task performance:

● **dorsal** (or **superior**): towards the top
● **ventral** (or **inferior**): towards the bottom
● **anterior** (or **rostral**): towards the front
● **posterior**: towards the back
● **lateral**: situated at the side
● **medial**: situated in the middle.

The German neurologist Korbinian Brodmann (1868–1918) produced a brain map based on differences in the distributions of cell types across cortical layers (see Figure 1.5).
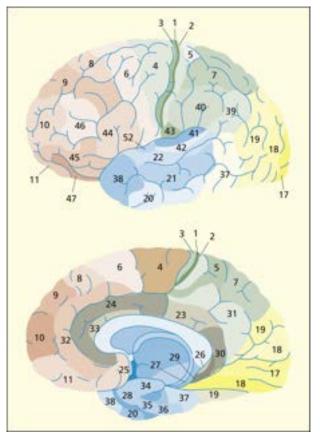

Figure 1.5
Brodmann brain areas on the lateral (top figure) and medial (bottom figure) surfaces.

He identified 52 areas. We will often refer to areas, for example, as BA17, which means Brodmann Area 17, rather than Brain Area 17!

Within cognitive neuroscience, brain areas are often described with reference to their main functions. For example, Brodmann Area 17 (BA17) is commonly called the primary visual cortex because it is strongly associated with the early processing of visual stimuli.
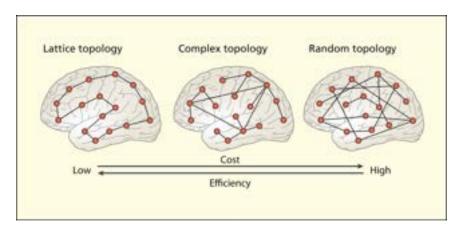
Brain organisation

In recent years, there has been considerable progress in identifying the connectome: this is a "wiring diagram" providing a complete map of the brain's neural connections. Why is it important to identify the connectome? First, as we will see, it advances our understanding of how the brain is organised. Second, identifying the brain's structural connections facilitates the task of understanding how it functions. More specifically, the brain's functioning is strongly constrained by its structural connections. Third, as we will see, we can understand some individual differences in cognitive functioning with reference to individual differences in the connectome.

Bullmore and Sporns (2012) used information about the connectome to address issues about brain organisation. They argued two major principles might determine its organisation. First, there is the *principle of cost control*: costs (e.g., use of energy and space) would be minimised if the brain consisted of limited, short-distance connections (see Figure 1.6). Second, there is the *principle of efficiency* (efficiency is the ability to integrate information across the brain). This can be achieved by having very numerous connections, many of which are long-distance (see Figure 1.6). These two principles are in conflict – you cannot have high efficiency at low cost.

You might imagine it would be best if our brains were organised primarily on the basis of efficiency. However, this would be incredibly costly – if all 100 billion brain neurons were interconnected, the brain would need to be 12½ miles wide (Ward, 2015)! In fact, neurons mostly connect with nearby neurons and no neuron is connected to more than about 10,000 other neurons. As a result, the human brain has a near-optimal trade-off between cost and efficiency (see Figure 1.6). Thus, our brains are reasonably efficient while incurring a manageable cost.

Figure 1.6
The left panel shows a brain network low in cost efficiency; the right panel shows a brain network high in cost efficiency; the middle panel shows the actual human brain in which there is moderate efficiency at moderate cost. Nodes are shown as orange circles.

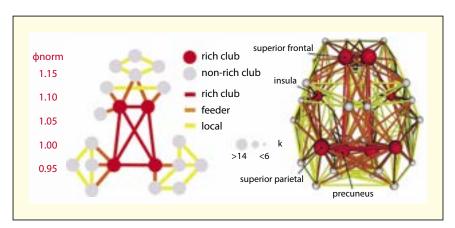From Bullmore and Sporns (2012). Reprinted with permission of Nature Reviews.

Figure 1.7
The organisation of the "rich club". It includes the precuneus, the superior frontal cortex, insular cortex, and superior parietal cortex. The figure also shows connections between rich club nodes, connections between non-rich club nodes (local connections), and connections between rich club and non-rich club nodes (feeder connections).

There is an important distinction between *modules* (small areas of tightly clustered connections) and *hubs* (regions having large numbers of connections to other regions). This is an efficient organisation as can be seen by analogy to the world's airports – the needs of passengers are best met by having numerous local airports (modules) and relatively few major hubs (e.g., Heathrow in London; Changi in Singapore; Los Angeles International Airport).

Collin et al. (2014) argued the brain's hubs are strongly interconnected and used the term "rich club" to refer to this state of affairs. The organisation of the rich club is shown in Figure 1.7: it includes the precuneus, the superior frontal cortex, insular cortex and superior parietal cortex. The figure also shows connections between rich club nodes, connections between non-rich club nodes (local connections), and connections between rich club and non-rich club nodes (feeder connections).

What light does a focus on brain network organisation shed on individual differences in cognitive ability? Hilger et al. (2017) distinguished between *global* efficiency (i.e., efficiency of the overall brain network) and *nodal* efficiency (i.e., efficiency of specific hubs or nodes). Intelligence was unrelated to global efficiency. However, it was positively associated with the efficiency of two hubs or nodes: the anterior insula and dorsal anterior cingulate cortex. The anterior insula is involved in the detection of task-relevant stimuli whereas the dorsal anterior cingulate cortex is involved in performance monitoring.

Techniques for studying brain activity: introduction

Technological advances mean we have numerous exciting ways of obtaining detailed information about the brain's functioning and structure. In principle, we can work out *where* and *when* specific cognitive processes occur in the brain. This allows us to determine the *order* in which different brain areas become active when someone performs a task. It also allows us to discover the extent to which two tasks involve the same brain areas.

Information concerning the main techniques for studying brain activity is contained in Table 1.2. Which technique is the best? There is no single (or simple) answer. Each technique has its own strengths and limitations,

KEY TERMS 🔑

Single-unit recording
An invasive technique for studying brain function, permitting the study of activity in single neurons.

Event-related potentials (ERPs)
The pattern of electroencephalograph (EEG) activity obtained by averaging the brain responses to the same stimulus (or very similar stimuli) presented repeatedly.

Positron emission tomography (PET)
A brain-scanning technique based on the detection of positrons; it has reasonable spatial resolution but poor temporal resolution.

Functional magnetic resonance imaging (fMRI)
A technique based on imaging blood oxygenation using an MRI machine; it provides information about the location and time course of brain processes.

Event-related functional magnetic resonance imaging (efMRI)
This is a form of **functional magnetic resonance imaging** in which patterns of brain activity associated with specific events (e.g., correct vs incorrect responses on a memory test) are compared.

Magneto-encephalography (MEG)
A non-invasive brain-scanning technique based on recording the magnetic fields generated by brain activity; it has good spatial and temporal resolution.

TABLE 1.2 MAJOR TECHNIQUES USED TO STUDY THE BRAIN

- **Single-unit recording**: This technique (also known as single-cell recording) involves inserting a micro-electrode 1/10,000th of a millimetre in diameter into the brain to study activity in single neurons. It is very sensitive: electrical charges of as little as one-millionth of a volt can be detected.

- **Event-related potentials (ERPs)**: The same stimulus (or very similar ones) are presented repeatedly, and the pattern of electrical brain activity recorded by several scalp electrodes is averaged to produce a single waveform. This technique allows us to work out the timing of various cognitive processes very precisely but its spatial resolution is poor.

- **Positron emission tomography (PET)**: This technique involves the detection of positrons (atomic particles emitted from some radioactive substances). PET has reasonable spatial resolution but poor temporal resolution and measures neural activity only indirectly.

- **Functional magnetic resonance imaging (fMRI)**: This technique involves imaging blood oxygenation using a magnetic resonance imaging (MRI) machine (described on p. 19). fMRI has superior spatial and temporal resolution to PET, but also provides an indirect measure of neural activity.

- **Event-related functional magnetic resonance imaging (efMRI)**: This "involves separating the elements of an experiment into discrete points in time, so that the cognitive processes (and associated brain responses) associated with each element can be analysed independently" (Huettel, 2012, p. 1152). Event-related fMRI is generally very informative and has become more popular recently.

- **Magneto-encephalography (MEG)**: This technique involves measuring the magnetic fields produced by electrical brain activity. It provides fairly detailed information at the millisecond level about the time course of cognitive processes, and its spatial resolution is reasonably good.

- **Transcranial magnetic stimulation (TMS)**: This is a technique in which a coil is placed close to the participant's head and a very brief pulse of current is run through it. This produces a short-lived magnetic field that generally (but not always) inhibits processing in the brain area affected. When the pulse is repeated several times in rapid succession, we have repetitive transcranial magnetic stimulation (rTMS). rTMS is used very widely.

  It has often been argued that TMS or rTMS causes a very brief "lesion". This technique has (jokingly!) been compared to hitting someone's brain with a hammer. More accurately, TMS often causes *interference* because the brain area to which it is applied is involved in task processing as well as the activity resulting from the TMS stimulation.

- **Transcranial direct current stimulation (tDCS)**: A weak electric current is passed through a given brain area for some time. The electric charge flows from a positive site (an anode) to a negative one (a cathode). Anodal tDCS increases cortical excitability and generally enhances performance. In contrast, cathodal tDCS decreases cortical excitability and mostly impairs performance.

and so experimenters match the technique to the research question. Of key importance, these techniques vary in the precision with which they identify the brain areas active when a task is performed (*spatial resolution*) and the time course of such activation (*temporal resolution*). Thus, they differ in their ability to provide precise information concerning *where* and *when* brain activity occurs.
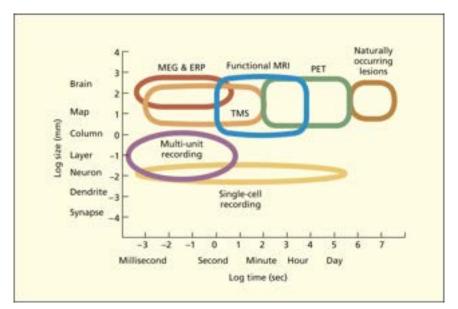
Figure 1.8
The spatial and temporal resolution of major techniques and methods used to study brain functioning.

From Ward (2006), adapted from Churchland & Sejnowski (1988).

Techniques for studying the brain: detailed analysis

We have introduced the main techniques for studying the brain. In what follows, we consider them in more detail.

Single-unit recording

The single-unit (or cell) recording technique is more fine-grain than any other technique (see Chapter 2). However, it is invasive and so rarely used with humans. An interesting exception is a study by Quiroga et al. (2005) on epileptic patients with implanted electrodes to identify the focus of seizure onset (see Chapter 3). A neuron in the medial temporal lobe responded strongly to pictures of Jennifer Aniston (the actor from *Friends*) but not to pictures of other famous people. We need to interpret this finding carefully. Only a tiny fraction of the neurons in that brain area were studied and it is highly improbable that none of the others would have responded to Jennifer Aniston.

Event-related potentials

Electroencephalography (EEG) is based on recordings of electrical brain activity measured at several locations on the surface of the scalp. Very small changes in electrical activity within the brain are picked up by scalp electrodes and can be seen on a computer screen. However, spontaneous or background brain activity can obscure the impact of stimulus processing on the EEG recording.

The answer to the above problem is to present the same stimulus (or very similar stimuli) many times. After that, the segment of the EEG following each stimulus is extracted and lined up with respect to the time of stimulus onset. These EEG segments are then averaged together to

> **KEY TERMS** 🔑
>
> **Transcranial magnetic stimulation (TMS)**
> A technique in which magnetic pulses briefly disrupt the functioning of a given brain area. It is often claimed that it creates a short-lived "lesion". More accurately, TMS causes interference when the brain area to which it is applied is involved in task processing as well as activity produced by the applied stimulation.
>
> **Transcranial direct current stimulation (tDCS)**
> A technique in which a very weak electrical current is passed through an area of the brain (often for several minutes); anodal tDCS often enhances performance, whereas cathodal tDCS often impairs it.
>
> **Electroencephalography (EEG)**
> Recording the brain's electrical potentials through a series of scalp electrodes.

produce a single waveform. This method produces event-related potentials from EEG recordings and allows us to distinguish the genuine effects of stimulation from background brain activity.

ERPs have excellent temporal resolution, often indicating when a given process occurred to within a few milliseconds. The ERP waveform consists of a series of positive (P) and negative (N) peaks, each described with reference to the time in milliseconds after stimulus onset. Thus, for example, N400 is a negative wave peaking at about 400 ms.

Behavioural measures (e.g., reaction times) typically provide only a *single* measure of time on each trial, whereas ERPs provide a *continuous* measure. However, ERPs do not indicate with precision *which* brain regions are most involved in processing, in part because skull and brain tissue distort the brain's electrical fields. In addition, ERPs are mainly of value when stimuli are simple and the task involves basic processes (e.g., target detection) triggered by task stimuli. Finally, we cannot study the most complex forms of cognition (e.g., problem solving) with ERPs because the processes by participants would typically change with increased practice.

Positron emission tomography (PET)

Positron emission tomography (PET) is based on the detection of positrons – atomic particles emitted by some radioactive substances. Radioactively labelled water (the tracer) is injected into the body and rapidly gathers in the brain's blood vessels. When part of the cortex becomes active, the labelled water moves there rapidly. A scanning device measures the positrons emitted from the radioactive water which leads to pictures of the activity levels in different brain regions.

PET has reasonable spatial resolution in that any active brain area can be located to within 5–10 mm. However, it has very poor temporal resolution – PET scans indicate the amount of activity in any given brain region over approximately 30 seconds. As a consequence, PET has now largely been superseded by fMRI (see p. 19).

The magnetic resonance imaging (MRI) scanner has proved an extremely valuable source of data in psychology.
Juice Images/Alamy Stock Photo.

Magnetic resonance imaging (MRI and fMRI)

Magnetic resonance imaging (MRI) involves using an MRI scanner containing a very large magnet (weighing up to 11 tons). A strong magnetic field causes an alignment of protons (subatomic particles) in the brain. A brief radio-frequency pulse is applied, which causes the aligned protons to spin and then regain their original orientations giving up a small amount of energy as they do so. The brightest regions in the MRI are those emitting the greatest energy. MRI scans can be obtained from numerous angles but tell us only about brain structure rather than its functions.

Happily, MRI can also be used to provide functional information in the form of functional magnetic resonance imaging (fMRI). Oxyhaemoglobin is converted into deoxyhaemoglobin when neurons consume oxygen, and deoxyhaemoglobin produces distortions in the local magnetic field (sorry this is so complex!). This distortion is assessed by fMRI and provides a measure of the concentration of deoxyhaemoglobin in the blood.

Technically, what is measured in fMRI is known as **BOLD** (blood oxygen-level-dependent contrast). Changes in the BOLD signal produced by increased neural activity take time, so the temporal resolution of fMRI is 2 or 3 seconds. However, its spatial resolution is very good (approximately 1 mm). Thus, fMRI has superior temporal and spatial resolution to PET.

Suppose we want to understand why people remember some items but not others. We can use event-related fMRI (efMRI), in which we consider

> **KEY TERMS** 🔑
>
> **BOLD**
> Blood oxygen-level-dependent contrast; this is the signal measured by **fMRI**.
>
> **Neural decoding**
> Using computer-based analyses of patterns of brain activity to work out which stimulus an individual is processing.

CAN COGNITIVE NEUROSCIENTISTS READ OUR BRAINS/MINDS? 🔍

There is much current interest in **neural decoding** – "determining what stimuli or mental states are represented by an observed pattern of neural activity" (Tong & Pratte, 2012, p. 483). This decoding involves complex computer-based analysis of individuals' patterns of brain activity and has sometimes been described as "brain reading" or "mind reading".

Kay et al. (2008) obtained impressive findings using neural decoding techniques. Two participants viewed 1,750 natural images and brain-activation patterns were obtained using fMRI. Computer-based approaches then analysed these patterns. After that, the participants were presented with 120 previously unseen natural images and fMRI data were collected. These fMRI data permitted correct identification of the image being viewed on 92% of the trials for one participant and 72% for the other. This is remarkable since chance performance was only 0.8%!

Huth et al. (2016) used more complex stimuli. They presented observers with clips taken from several movies including *Star Trek* and *Pink Panther 2* while using fMRI. Decoding accuracy was reasonably successful in identifying general object categories (e.g., *animal*), specific object categories (e.g., *canine*) and various actions (e.g., *talk*; *run*) presented in the movie clips.

Research on neural decoding can enhance our understanding of human visual perception. However, successful decoding of an object using the pattern of brain activation in a given brain region does *not* necessarily mean that region is causally involved in observers' identification of that object. Several reasons why we need to be cautious when interpreting findings from neural decoding studies are discussed by Popov et al. (2018). For example, some aspects of brain activity in response to visual stimuli are *irrelevant* to the observer's perceptual representation. In an experiment, computer analysis of brain activity in macaques successfully classified various stimuli presented to them that the macaques themselves could not distinguish (Hung et al., 2005).

each participant's patterns of brain activation for remembered and forgotten items. Wagner et al. (1998) recorded fMRI while participants learned a list of words. There was more brain activity during learning for words subsequently recognised than those subsequently forgotten. These findings suggest forgotten words were processed less thoroughly than remembered words during learning.

Evaluation

fMRI is the dominant technique within cognitive neuroscience. Its value has increased over the years with the introduction of more powerful MRI scanners. Initially, most scanners had a field strength of 1.5 T but recently scanners with field strengths of up to 7 T have become available. As a result, submillimetre spatial resolution is now possible (Turner, 2016).

What are fMRI's limitations? The main ones are as follows:

(1) It has relatively poor temporal resolution. As a result, it is uninformative about the order in which different brain regions (and cognitive processes) are used during performance of a task. However, other techniques within cognitive neuroscience can be used in conjunction with fMRI in order to achieve good spatial and temporal resolution.
(2) It provides an *indirect* measure of underlying neural activity. "The BOLD signal primarily measures the input and processing of neural information . . . but not the output signal transmitted to other brain regions" (Shifferman, 2015, p. 60).
(3) Various complex processes are used by researchers to take account of the fact that all brains differ. This involves researchers changing the raw fMRI data and poses the danger that "BOLD-fMRI neuroimages represent mathematical constructs rather than physiological reality" (Shifferman, 2015).
(4) There are important constraints on the visual stimuli that can be presented to participants when lying in the scanner and on the types of responses they can make. There can be particular problems with auditory stimuli because the scanner is noisy.

Magneto-encephalography

The electric currents that the brain generates are associated with a magnetic field. This magnetic field is assessed by magneto-encephalography (MEG) involving at least 200 devices on the scalp. This technique has very good spatial and temporal resolution. However, it is extremely expensive and this has limited its use.

Transcranial magnetic stimulation

Transcranial magnetic stimulation (TMS) is a technique in which a coil (often in the shape of a figure of eight) is placed close to the participant's head. A very brief (under 1 ms) but large magnetic pulse of current is run

through it. This creates a short-lived magnetic field generally leading to inhibited processing in the directly affected area (about 1 cc in extent). More specifically, the magnetic field created leads to electrical stimulation in the brain. In practice, several magnetic pulses are typically given in rapid succession – this is repetitive transcranial magnetic stimulation (rTMS). Most research has used rTMS but we will often simply use the more general term TMS.

Transcranial magnetic stimulation coil.
University of Durham/Simon Fraser/Science Photo Library.

What is an appropriate control condition against which to compare the effects of TMS or rTMS? We could compare task performance with and without TMS. However, TMS creates a loud noise and muscle twitching at the side of the forehead and these effects might lead to impaired performance. Applying TMS to a non-critical brain area (irrelevant for task performance) often provides a suitable control condition. It is typically predicted task performance will be worse when TMS is applied to a critical area rather than a non-critical one because it produces a temporary "lesion" or interference to the area targeted.

Evaluation

What are TMS's strengths? First, it permits causal inferences – if TMS applied to a particular brain area impairs task performance, we can infer that brain area is necessary for task performance. Conversely, if TMS has no effects on task performance, we can conclude the brain area affected by it is irrelevant. In that respect, it resembles cognitive neuropsychology.

Second, TMS research is more flexible than cognitive neuropsychology. For example, we can compare any given individual's performance with and without a "lesion" with TMS. This is rarely possible with brain-damaged patients.

Third, TMS research is also more flexible than cognitive neuropsychology because the researcher controls the brain area(s) affected. In addition, the temporary "lesions" created by TMS typically cover a smaller brain area than patients' lesions. This is important because the smaller the brain affected, the easier it generally is to interpret task-performance findings.

Fourth, with TMS research, we can ascertain *when* a brain area is most activated. For example, the presentation of a visual stimulus leads to processing proceeding rapidly to higher visual levels (feedforward processing). According to Lamme (2010), conscious visual perception typically requires subsequent recurrent processing proceeding in the opposite direction from higher levels to lower ones. Koivisto et al. (2011) used TMS to disrupt recurrent processing. As predicted, this impaired conscious visual perception.

What are the limitations of TMS research? First, the effects of TMS are complex and not fully understood. When TMS was first introduced, it was assumed it would disrupt performance. That is, the most common finding. However, Luber and Lisanby (2014) reviewed 61 studies in which performance speed and/or accuracy was *enhanced*!

How can TMS enhance performance? It sometimes increases neural activity in areas adjacent to the one stimulated and so increases their processing efficiency (Luber & Lisanby, 2014). There is also much evidence for compensatory flexibility (Hartwigsen, 2018) – disruption to cognitive processing within a given brain area caused by TMS is *compensated for* by the recruitment of other brain areas.

Second, it is hard to establish the precise brain areas affected by TMS. For example, adverse effects of TMS on performance might occur because it interferes with communication between two brain areas at some distance from the stimulation point. This issue can be addressed by combining TMS with neuroimaging techniques to clarify its effects (Valero-Cabré et al., 2017).

Third, TMS can only be applied to brain areas lying beneath the skull but with overlying muscle. That limits its overall usefulness.

Fourth, there are safety issues with TMS. It has very occasionally caused seizures in participants despite stringent rules designed to ensure their safety.

Transcranial direct current stimulation (tDCS)

As mentioned earlier, anodal tDCS *increases* cortical excitability whereas cathodal tDCS *decreases* cortical excitability. The temporal and spatial resolution of tDCS is less than TMS (Stagg & Nitsche, 2011). However, anodal tDCS has a significant advantage over TMS in that it often facilitates or enhances cognitive functioning. As a result, anodal tDCS is increasingly used to reduce adverse effects of brain damage on cognitive functioning (Stagg et al., 2018). Some of these beneficial effects on cognition are relatively long-lasting. Another advantage of tDCS is that it typically causes little or no discomfort.

Much progress has been made in understanding the complex mechanisms associated with tDCS. However, "Knowledge about the physiological effects of tDCS is still not complete" (Stagg et al., 2018, p. 144).

Overall strengths

Cognitive neuroscience has contributed substantially to our understanding of human cognition. We discuss supporting evidence for that statement throughout the book in areas that include perception, attention, learning, memory, language comprehension, language production, problem solving, reasoning, decision-making and consciousness. Here we identify its major strengths.

First, cognitive neuroscience has helped to resolve theoretical controversies and issues that had proved intractable with purely behavioural studies (Mather et al., 2013). The main reason is that cognitive neuroscience adds considerably to the information available to researchers (Poldrack & Yarkoni, 2016). Below we briefly consider two examples:

(1) Listeners hearing degraded speech find it much more intelligible when it is accompanied by visually presented words matching (rather than not matching) the auditory input. There has been much theoretical

controversy concerning *when* this visually presented information influences speech perception (see Chapter 9). Does it occur *early* and so directly influence basic auditory processes or does it occur *late* (after basic auditory processing has finished)? Wild et al. (2012) found there was more activity in brain areas involved in early auditory processing when the visual input matched the auditory input. This strongly suggests visual information directly influences basic auditory processes.

(2) There has been much theoretical controversy as to whether visual imagery resembles visual perception (see Chapter 3). Behavioural evidence has proved inconclusive. However, neuroimaging has shown two-thirds of the brain areas activated during visual perception are also activated during visual imagery (Kosslyn, 2005). Even brain areas involved in the early stages of visual perception are often activated during visual imagery tasks (Kosslyn & Thompson, 2003). Thus, there are important similarities. Functional neuroimaging has also revealed important differences between the processes involved in visual perception and imagery (Dijkstra et al., 2017b).

> **KEY TERM** 🔑
>
> **Functional specialisation**
> The assumption that each brain area or region is specialised for a specific function (e.g., colour processing; face processing).

Second, the incredible richness of neuroimaging data means cognitive neuroscientists can (at least in principle) construct theoretical models accurately mimicking the complexities of brain functioning. In contrast, cognitive neuropsychology, for example, is less flexible and more committed to the notion of a modular brain organisation.

Third, over 10,000 fMRI studies within cognitive neuroscience have been published. Many meta-analyses based on these studies have been carried out to understand brain-cognition relationships (Poldrack & Yarkoni, 2016). Such meta-analyses "provide highly robust estimates of the neural correlates of relatively specific cognitive tasks" (p. 592). At present, this approach is limited because we do not know the pattern of activation associated with any given cognitive process – data are coded with respect to particular tasks rather than underlying cognitive processes.

Fourth, neuroimaging data can often be *re-analysed* based on theoretical developments. For example, early neuroimaging research on face processing suggested it occurs mostly within the fusiform face area (see Chapter 3). However, the assumption that face processing involves a network of brain regions provides a more accurate account (Grill-Spector et al., 2017). Thus, cognitive neuroscience can be self-correcting.

More generally, cognitive neuroscience has shown the assumption of functional specialisation (each brain area is specialised for a different function) is oversimplified. We can contrast the notions of functional specialisation and functional integration (positive correlations of various brain areas within a network). For example, conscious perception depends on coordinated activity across several brain regions and so involves functional integration (see Chapter 16).

Overall limitations

We turn now to general issues raised by cognitive neuroscience. We emphasise fMRI research because that technique has been used most often.

KEY TERM 🔑

Reverse inference
As applied to functional neuroimaging, it involves arguing backwards from a pattern of brain activation to the presence of a given cognitive process.

First, many cognitive neuroscientists *over-interpret* their findings by assuming one-to-one links between cognitive processes and brain areas. For example, activation in a particular small brain region (a "blob") is interpreted as being the "love area" and another small region was interpreted as being the "face processing area". This approach has been described (unflatteringly) as "blobology".

Blobology is in decline. However, there is still undue reliance on reverse inference – the involvement of a given cognitive process is inferred from activation within a given brain region. For example, face recognition is typically associated with activation within the fusiform face area, which led many researchers to identify that area as specifically involved in face processing. This is incorrect in two ways (see Chapter 3): (1) the fusiform face area is activated in response to many different kinds of objects as well as faces (Downing et al., 2006); and (2) several other brain areas (e.g., occipital face area) are also activated during face processing.

Second, cognitive neuroscience is rarely used to test cognitive theories. For example, Tressoldi et al. (2012) reviewed 199 studies published in 8 journals. They found 89% of these studies focused on localising the brain areas associated with brain processes and only 11% tested a theory of cognition.

There is some validity to Tressoldi et al.'s (2012) argument. However, it has become less persuasive because of the rapidly increasing emphasis within cognitive neuroscience on theory testing.

Third, it is hard to bridge the divide between psychological processes and concepts and patterns of brain activation. As Harley (2012) pointed out, we may never find brain patterns corresponding closely to psychological processes such as "attention" or "planning". Harley (2012, p. 1372) concluded as follows: "Our language and thought may not divide up in the way in which the brain implements these processes."

Fourth, it is sometimes hard to replicate findings within cognitive neuroscience. For example, Uttal (2012) compared the findings from different brain-imaging meta-analyses on a given cognitive function. More specifically, he identified the Brodmann areas associated with a given cognitive function in two meta-analyses. Uttal then worked out the Brodmann areas activated in both meta-analyses and divided this by the total number of Brodmann areas identified in at least one meta-analysis. If the two meta-analyses were in total agreement, the resultant figure would be 100%. The actual figure varied between 14% and 51%!

Uttal's (2012) findings seem devastating – after all, we might expect meta-analyses based on numerous studies to provide very reliable evidence. However, many apparent discrepancies occurred mainly because one meta-analysis reported activation in fewer brain areas than the other. This often happened because of stricter criteria for deciding a given brain area was activated (Klein, 2014).

Fifth, false-positive findings (i.e., mistakenly concluding that random activity in a given brain area is task-relevant) are common (Yarkoni et al., 2010; see discussion later on p. 25). There are several reasons for this. For example, researchers have numerous options when deciding precisely how to analyse their fMRI data (Poldrack et al., 2017). In addition, most neuroimaging studies produce huge amounts of data and researchers

sometimes fail to adjust required significance levels appropriately.

Bennett et al. (2009) provided an example of a false-positive finding. They asked their participant to determine the emotions shown in photographs. When they did not adjust required significance levels, there was significant evidence of brain activation (see Figure 1.9). Amusingly, the participant was a dead salmon so we can be certain the "finding" was a false positive.

Figure 1.9

Areas showing greater activation in a dead salmon when presented with photographs of people than when at rest.
From Bennett et al. (2009). With kind permission of the authors.

Sixth, most brain-imaging techniques reveal only *associations* between patterns of brain activation and behaviour. Such associations are purely correlational and do not establish the brain regions activated are *necessary* for task performance. For example, brain activation might also be caused by participants engaging in unnecessary monitoring of their performance or attending to non-task stimuli.

Seventh, many cognitive neuroscientists previously assumed most brain activity is driven by environmental or task demands. As a result, we might expect relatively large increases in brain activity in response to such demands. That is *not* the case. The increased brain activity when someone performs a cognitive task typically adds less than 5% to resting brain activity. Surprisingly several brain areas exhibit *decreased* activity when a cognitive task is performed. Of importance here is the **default mode network** (Raichle, 2015). It consists of an interconnected set of brain regions (including the ventral medial and dorsal medial prefrontal cortex, and the posterior cingulate cortex) more active during rest than during performance of a task. Its functions include mind wandering, worrying and daydreaming.

The key point is that patterns of brain activity in response to any given cognitive task reflect *increased* activity associated with task processing and *decreased* activity associated with reduced activity within the default mode network. Such complexities complicate the interpretation of neuroimaging data.

Eighth, cognitive neuroscience shares with cognitive psychology problems of ecological validity (applicability to everyday life) and paradigm specificity (findings not generalising across paradigms). Indeed, the problem of ecological validity may be greater in cognitive neuroscience. For example, participants in fMRI studies lie on their backs in claustrophobic and noisy conditions and have very restricted movement – not much like everyday life! Gutchess and Park (2006) found recognition memory was significantly worse in an MRI scanner than in an ordinary laboratory. Presumably the scanner provided a more distracting or anxiety-creating environment.

Ninth, we must avoid what Ali et al. (2014) termed "neuroenchantment" – exaggerating the importance of neuroimaging to our understanding of cognition. Ali et al. provided a striking example of neuroenchantment. College students were exposed to a crudely built mock brain scanner (including a discarded hair dryer!) (see Figure 1.10). They were asked to think about
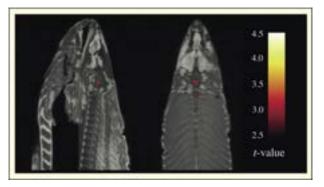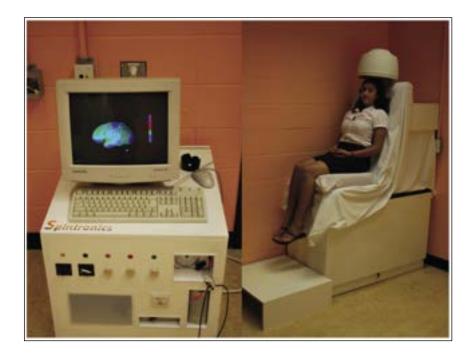
KEY TERM

Default mode network
A network of brain regions that is active "by default" when an individual is not involved in a current task; it is associated with internal processes including mind wandering, remembering the past and imagining the future.

Figure 1.10
The primitive mock neuroimaging device used by Ali et al. (2014).

the answers to various questions (e.g., name a country). The mock neuro-imaging device apparently "read their minds" and worked out exactly what they were thinking. Amazingly, three-quarters of the student participants believed this was genuine rather than being due to the researcher's trickery?

COMPUTATIONAL COGNITIVE SCIENCE

There is an important distinction between computational modelling and artificial intelligence. Computational modelling involves programming computers to model or mimic human cognitive functioning. Thus, cognitive modellers "have the goal of understanding the human mind through computer simulation" (Taatgen et al., 2016, p. 1). In contrast, artificial intelligence involves constructing computer systems producing intelligent outcomes but typically in ways different from humans. Consider Deep Blue, the IBM computer that defeated the world chess champion Garry Kasparov on 11 May 1997. Deep Blue processed up to 200 million positions per second, which is vastly more than human chess players (see Chapter 12).

The IBM computer Watson also shows the power of artificial intelligence. This computer competed on the American quiz show *Jeopardy* against two of the most successful human contestants ever on that show: Brad Rutter and Ken Jennings. The competition took place between 14 and 16 February 2011, and Watson won the $1 million first prize. Watson had the advantage over Rutter and Jennings of having access to 10 million documents (200 million pages of content). However, Watson had the disadvantage of being less sensitive to subtleties contained in the questions.

In the past (and even nowadays), many experimental cognitive psychologists expressed their theories in vague verbal statements (e.g., "Information

KEY TERMS 🔑

Computational modelling
This involves constructing computer programs that simulate or mimic human cognitive processes.

Artificial intelligence
This involves developing computer programs that produce intelligent outcomes.

The IBM Watson and two human contestants (Ken Jennings and Brad Rutter).

Ben Hider/Getty Images.

from short-term memory is transferred to long-term memory"). This made it hard to provide precise predictions from the theory and to decide whether the evidence fitted the theory. As Murphy (2011) pointed out, verbal theories provided theorists with undesirable "wiggle room". In contrast, a computational model "requires the researchers to be explicit about a theory in a way that a verbal theory does not" (Murphy, 2011, p. 300). Implementing a theory as a program is a good way to check it contains no hidden assumptions or imprecise terms. This often reveals that the theory makes predictions the theorist had not realised!

There are issues concerning the relationship between the performance of a computer program and human performance (Costello & Keane, 2000). For example, a program's speed doing a simulated task can be affected by psychologically irrelevant features (e.g., the power of the computer). Nevertheless, the various materials presented to the program should result in differences in program operation time correlating closely with differences in participants' reaction times with the same materials.

Types of models

Most computational models focus on specific aspects of human cognition. For example, there are successful computational models providing accounts of reading words and non-words aloud (Coltheart et al., 2001; Perry et al., 2007, 2014; Plaut et al., 1996) (see Chapter 9).

More ambitious computational models provide **cognitive architectures** – "models of the fixed structure of the mind" (Rosenbloom et al., 2017, p. 2). Approximately 300 cognitive architectures have been proposed over the years (Kotseruba & Tsotsos, 2018).

Note that a cognitive architecture typically has to be supplemented with the knowledge required to perform a given task to produce a fully fledged computational model (Byrne, 2012). Anderson et al. (2004) proposed an

> **KEY TERM** 🔑
>
> **Cognitive architecture**
> Comprehensive framework for understanding human cognition in the form of a computer program.

KEY TERMS 🗝

Connectionist models
Models in computational cognitive science consisting of interconnected networks of simple units or **nodes**; the networks exhibit learning through experience and specific items of knowledge are distributed across numerous units.

Neural network models
Computational models in which processing involves the simultaneous activation of numerous interconnected **nodes** (basic units).

Nodes
The basic units within a **neural network model**.

Back-propagation
A learning mechanism in connectionist models based on comparing actual responses to correct ones.

especially influential cognitive architecture in their Adaptive Control of Thought-Rational (ACT-R) (discussed on p. 30).

Connectionism

Connectionist models (also called **neural network models**) typically consist of interconnected networks of simple units (or **nodes**) that exhibit learning.

Connectionist or neural network models use elementary units or nodes connected together in structures or layers (see Figure 1.11). First, a layer of input nodes codes the input. Second, activation caused by input coding spreads to a layer of hidden nodes. Third, activation spreads to a layer of output nodes.

Of major importance, the basic model shown in Figure 1.11 can learn simple additions. If input node 1 is active, output node 1 will also become active. If input nodes 1 and 2 are active, output node 3 will become active. If *all* input nodes are active, output node 10 will become active and so on. The model compares the *actual* output against the *correct* output. If there is a *discrepancy*, the model learns to adjust the weights of the connections between the nodes to produce the correct output. This is known as backward propagation of errors or **back-propagation**, and it allows the model to learn the appropriate responses without being explicitly programmed to do so.

Numerous connectionist models have been constructed using the basic architecture shown in Figure 1.11. Recent connectionist models often have several hidden layers (they are called deep neural networks). Connectionist models involve distributed representations whereas other computational models involve localist representations (Bowers, 2017a). In the former case, each node or unit responds to multiple stimulus categories and so a given word or object is represented by the pattern of activation across many nodes or units.

In the latter case, each node or unit responds most actively to a single meaningful stimulus category (e.g., a given word or object).

Interest in the connectionist approach was triggered initially by Rumelhart et al. (1986) and McClelland et al. (1986) with their parallel distributed processing models. These models (like most models based on distributed representations) exhibit learning. Other influential connectionist models are Plaut et al.'s (1996) reading model (Chapter 9) and McClelland & Elma's (1986) TRACE model of spoken word recognition (Chapter 9).

In contrast, computational models based on localist representations often do not exhibit learning because the representations contain all the required information. Examples of such models in this book include Coltheart et al.'s (2001) reading model (Chapter 9) and the speech production models of Dell (1986) and Levelt et al. (1999; see Chapter 11).

It has often been assumed localist models are biologically implausible because they seem to imply a *single* neuron responds to stimuli
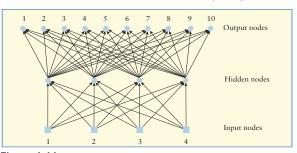
Figure 1.11
Architecture of a basic three-layer connectionist network.

from a given category. However, many neurons show considerable selectivity because they respond to only a very small fraction of stimuli. For example, consider a study by Quiroga et al. (2005) mentioned earlier. They presented epileptic patients with 100 images including famous individuals and landmark buildings. On average, responsive neurons responded to only approximately 3% of the images. The issues are complex. However, it is not clear localist models are less biologically plausible than distributed models (Bowers, 2017b).

Evaluation

There are many successful connectionist or neural network models (some are mentioned above). Such models (discussed at various points in this book) have provided valuable insights into human cognition and the models have become increasingly sophisticated over time. A general strength of neural networks is that they "exhibit robust flexibility in the face of the challenges posed by the real world" (Garson, 2016, p. 3). That is one reason why neural networks can perform numerous different kinds of cognitive tasks. Finally, there are intriguing similarities between the brain, with its numerous units (neurons) and synaptic connections, and neural networks, with their units (nodes) and connections.

What are the limitations of connectionist models? First, there is an issue with the common assumption that connectionist models are distributed. If two words are presented at the same time, this can lead to *superimposing* two patterns over the same units or nodes making it hard (or impossible) to decide which activated units or nodes belong to which word. This causes superposition catastrophe (Bowers, 2017a).

Second, there are many examples of neural networks that can make associations and match patterns. However, it has proved much harder to develop neural networks that can learn general rules (Garson, 2016).

Third, the analogy between neural networks and the brain is very limited. In essence, the latter is hugely more complex than the former.

Fourth, back-propagation implies learning will be slow, whereas humans sometimes exhibit one-trial learning (Garson, 2016). Furthermore, there is little or no evidence of back-propagation in the human brain (Mayor et al., 2014).

Production systems

Production systems consist of numerous "IF . . . THEN" production rules. Production rules can take many forms. However, an everyday example is: "If the green man is lit up, then cross the road." There is also a working memory (i.e., a system holding information currently being processed). If information from the environment that "green man is lit up" reaches working memory, it will match the IF part of the rule in long-term memory and trigger the THEN part of the rule (i.e., cross the road).

Production systems vary but generally have the following characteristics:

- numerous IF . . . THEN rules;
- a working memory containing information;

KEY TERMS

Production systems
These consist of very large numbers of "IF . . . THEN" production rules and a working memory containing information.

Production rules
"IF . . . THEN" or condition-action rules in which the action is carried out whenever the appropriate condition is present.

Working memory
A limited-capacity system used in the processing and brief holding of information.

- a production system that operates by matching the contents of working memory against the IF parts of the rules and then executing the THEN parts;
- if information in working memory matches the IF parts of two rules, a conflict-resolution strategy selects one.

Adaptive Control of Thought-Rational (ACT-R) and beyond

As mentioned earlier, Anderson et al. (2004) proposed ACT-R, which was subsequently developed (e.g., Anderson et al., 2008). ACT-R assumes the cognitive system consists of several modules (relatively independent subsystems). It combines computational cognitive science with cognitive neuroscience by identifying the brain regions associated with each module (see Figure 1.12). Four modules are of special importance:

(1) *Retrieval module*: it maintains the retrieval cues needed to access information; its proposed location is the inferior ventrolateral prefrontal cortex.
(2) *Imaginal module*: it transforms problem representations to assist in problem solving; it is located in the posterior parietal cortex.
(3) *Goal module*: it keeps tracks of an individual's intentions and controls information processing; it is located in the anterior cingulate cortex.
(4) *Procedural module*: it uses production (IF . . . THEN) rules to determine what action will be taken next; it is located at the head of the caudate nucleus within the basal ganglia.

Each module has a buffer associated with it containing a limited amount of important information. *How* is information from these buffers integrated? According to Anderson et al. (2004, p. 1058): "A central production system can detect patterns in these buffers and take co-ordinated action."
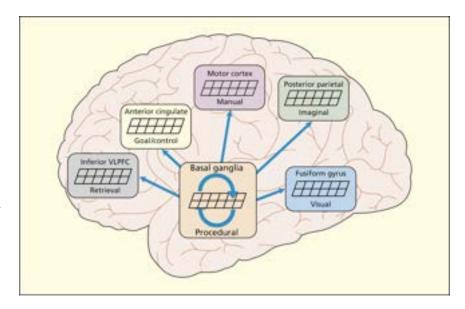
Figure 1.12
The main modules of the ACT-R (Adaptive Control of Thought-Rational) cognitive architecture with their locations within the brain.

Reprinted from Anderson et al. (2008). Reprinted with permission of Elsevier.

If several productions could be triggered by the information contained in the buffers, one is selected based on the value or gain associated with each outcome plus the amount of time or cost incurred in achieving that outcome.

ACT-R represents an impressive attempt to provide a theoretical framework for understanding information processing and performance on numerous cognitive tasks. It is also impressive in seeking to integrate computational cognitive science with cognitive neuroscience.

What are ACT-R's limitations? First, it is very hard to test such a wide-ranging theory. Second, areas of prefrontal cortex (e.g., dorsolateral prefrontal cortex) generally assumed to be of major importance in cognition are de-emphasised. Third, as discussed earlier, research within cognitive neuroscience increasingly reveals the importance to cognitive processing of brain networks rather than specific regions. Fourth, in common with most other cognitive architectures, ACT-R has a knowledge base that is substantially smaller than that possessed by humans (Lieto et al., 2018). This reduces the applicability of ACT-R to human cognitive performance.

Standard model of the mind

Dozens of cognitive architectures have been proposed and it is difficult to compare them. Laird et al. (2017; see also Rosenbloom et al., 2017) recently proposed a standard model emphasising *commonalities* among major cognitive architectures including ACT-R (see Figure 1.13).

Figure 1.13 may look unimpressive because it represents the model at a very general level. However, the model contains numerous additional assumptions (Laird et al., 2017). First, procedural memory has special importance because it has access to the whole of working memory; in contrast, the other modules have access only to specific aspects of working memory.
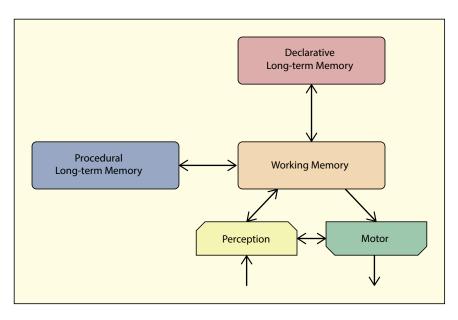

Figure 1.13
The basic structure of the standard model involving five independent modules. Declarative memory (see Glossary) stores facts and events whereas procedural memory (see Glossary) stores knowledge about actions.

Second, the crucial assumption is that there is a *cognitive cycle* lasting approximately 50 ms per cycle. What happens is that: "Procedural memory induces the selection of a single deliberate act per cycle, which can modify working memory, initiate the retrieval of knowledge from long-term declarative memory, initiate motor actions . . ., and provide top-down influence to perception" (Laird et al., 2017, p. 3). The cognitive cycle involves serial processing but parallel processing can occur within any module as well as between them.

The standard model is useful but incomplete. For example, it does not distinguish between different types of declarative memory (e.g., episodic and semantic memory; see Chapter 7). In addition, it does not account for emotional influences on cognitive processing.

Links with other approaches

ACT-R represents an impressive attempt to apply computational models to cognitive neuroscience. There has also been interest in applying such models to data from brain-damaged patients. Typically, the starting point is to develop a computational model accounting for the performance of healthy individuals on some task. After that, aspects of the computational model or program are altered to simulate "lesions", and the effects on task performance are assessed. Finally, the lesioned model's performance is compared against that of brain-damaged patients (Dell & Caramazza, 2008).

Overall strengths

Computational cognitive science has several strengths. First, the development of cognitive architectures can provide an overarching framework for understanding the cognitive system. This would be a valuable achievement given that much research in cognitive psychology is limited in scope and suffers from paradigm specificity. Laird et al.'s (2017) standard model represents an important step on the way to that achievement.

Second, the scope of computational cognitive science has increased. Initially, it was applied mainly to behavioural data. More recently, however, it has been applied to functional neuroimaging data (e.g., Anderson et al., 2004) and EEG data (Anderson et al., 2016a). Why is this important? As Taatgen et al. (2016, p. 3) pointed out: "The link to neuroimaging is critical in establishing that the hypothesised processing steps in cognitive models have plausibility in reality."

Third, rigorous thinking is required to develop computational models because computer programs must contain detailed information about the processes involved in performing any given task. In contrast, many theories within traditional cognitive psychology are vaguely expressed and the predictions following from their assumptions are unclear.

Fourth, progress is increasingly made by using nested incremental modelling. In essence, a new model builds on the strengths of previous related models while eliminating (or reducing) their weaknesses and accounting for additional data. For example, Perry et al. (2007; Chapter 9) put forward a connectionist dual-process model (CDP+) of reading aloud that improved on the model on which it was based.

Overall limitations

What are the main limitations of the computational cognitive science approach? First, there is Bonini's paradox: as models become more accurate and complete, they can become as hard to understand as the complex phenomena they are designed to explain. Conversely, models easy to understand are typically inaccurate and incomplete. Many computational modellers have responded to this paradox by focusing on the essence of the phenomena and ignoring the minor details (Milkowski, 2016, p. 1459).

Second, many computational models are hard to falsify. The ingenuity of computational modellers means many models can account for numerous behavioural findings (Taatgen et al., 2016). This issue can be addressed by requiring computational models to explain neuroimaging findings as well as behavioural ones.

Third, some computational models are less successful than they appear. One reason is *overfitting* in which a model accounts for noise in the data as well as genuine effects (Ziegler et al., 2010). Overfitting often means a model seems to account very well for a given data set but is poor at predicting new data (Yarkoni & Westfall, 2017).

Fourth, most computational models ignore motivational and emotional factors. Norman (1980) distinguished between a cognitive system (the Pure Cognitive System) and a biological system (the Regulatory System). Computational cognitive science typically de-emphasises the Regulatory System even though it often strongly influences the Pure Cognitive System. This issue can be addressed by developing computational models indicating how emotions modulate cognitive processes (Rodriguez et al., 2016).

Fifth, many computational models are very hard to understand (Taatgen et al., 2016). Why is this so? Addyman and French (2012, p. 332) identified several reasons:

> Everyone still programs in [their] own favourite programming language, source code is rarely made available, . . . even for other modellers, the profusion of source code in a multitude of programming languages, writing without programming guidelines, makes it almost impossible to access, check, explore, re-use or continue to develop [models].

Computational modellers often fail to share their source codes and models because they have perceived ownership over their own research and are concerned about losing control over it (Fecher et al., 2015).

COMPARISONS OF MAJOR APPROACHES

We have discussed the major approaches to human cognition at length and you may wonder which is the most useful and informative. However, that is *not* the best way of thinking about the issues for various reasons:

(1) An increasing amount of research involves two or more different approaches.

(2) Each approach makes its own distinctive contribution and so all are required. By analogy, it would be pointless asking whether a driver is more or less useful than a putter for a golfer – both are essential.

(3) Each approach has its own limitations as well as strengths (see Table 1.2). The optimal solution in such circumstances is to use **converging operations** – several different research methods are used to address a given theoretical issue with the strengths of one method balancing out the limitations of other methods. If different methods produce the same answer, that provides stronger evidence than could be obtained using a single method. If different methods produce different answers, further research is required to clarify matters. Note that using convergent operations is more difficult and demanding than using a single approach (Brase, 2014).

In writing this book, our coverage of each topic emphasises research most enhancing our understanding. As a result, any given approach (e.g., cognitive neuroscience; cognitive neuropsychology) is strongly represented when we discuss some topics but is much less well represented with other topics.

IS THERE A REPLICATION CRISIS?

Replication (the ability to repeat the findings of previous research using the same or similar experimental methods) is of central importance to psychology (including cognitive psychology). In recent years, however, there have been concerns about the extent to which findings can be replicated, leading Shrout and Rodgers (2018, p. 487) to refer to "the replication crisis" in psychology.

An important trigger for these concerns was an influential article in which the replicability of 100 studies published in leading psychology journals was assessed (Open Science Collaboration, 2015). Only 36% of the findings reported in these studies were replicated. Within cognitive psychology, only 21 out of 42 findings (50%) were replicated.

Only 50% of findings were reproduced! This is perhaps less problematical than it sounds. The complexities of research mean individual studies can provide only an estimate of the "true" state of affairs rather than definitive evidence (Stanley & Spence, 2014).

Why are there problems with replicating findings in cognitive psychology (and psychology generally)? A major reason is the sheer complexity of human cognition – cognitive processing and performance are influenced by numerous factors (many of which are not controlled or manipulated). As a result, even replicated findings often differ considerably in terms of the size of the effects obtained (Stanley et al., 2018).

Another reason is that experimenters sometimes use questionable research practices *exaggerating* the true statistical significance of their data. One example is p-hacking (selective reporting), in which "Researchers conduct many analyses on the same data set and just report those that are statistically significant" (Simmons et al., 2018, p. 255). Another example involves researchers proposing hypotheses *after* research results are known rather than *before* as should be the case (Shrout & Rodgers, 2018).

TABLE 1.3 STRENGTHS AND LIMITATIONS OF MAJOR APPROACHES TO HUMAN COGNITION

Strengths	Limitations
Experimental cognitive psychology	
1. The first systematic approach to understanding human cognition	1. Most cognitive tasks are complex and involve many different processes
2. The source of most theories and tasks used by the other approaches	2. Behavioural evidence provides indirect evidence concerning internal processes
3. It is enormously flexible and can be applied to any aspect of cognition	3. Theories are sometimes vague and hard to test empirically.
4. It has produced numerous important replicated findings	4. Findings sometimes do not generalise because of paradigm specificity
5. It has strongly influenced social, clinical and developmental psychology	5. There is a lack of an overarching theoretical framework
Cognitive neuropsychology	
1. Double dissociations have provided strong evidence for various processing modules	1. Patients may develop compensatory major strategies not found in healthy individuals
2. Causal links can be shown between brain damage and cognitive performance	2. Most of its theoretical assumptions (e.g., the mind is modular) seem too extreme
3. It has revealed unexpected complexities in cognition (e.g., language)	3. Detailed cognitive processes and their interconnectedness are often not specified
4. It transformed memory and language research	4. There has been excessive reliance on single-case studies
5. It straddles the divide between cognitive psychology and cognitive neuroscience	5. Brain plasticity complicates interpreting findings
Cognitive neuroscience: functional neuroimaging + ERPs + TMS	
1. Great variety of techniques offering excellent temporal or spatial resolution	1. Functional neuroimaging techniques provide essentially correlational data
2. Functional specialisation *and* brain integration can be studied	2. Much over-interpretation of data involving reverse inferences
3. TMS is flexible and permits causal inferences	3. There are many false positives and replication failures
4. Rich data permit assessment of integrated brain processing as well as specialised functioning	4. It has generated very few new theories
5. Resolution of complex theoretical issues	5. Difficulty in relating brain activity to psychological processes
Computational cognitive science	
1. Theoretical assumptions are spelled out with precision	1. Many computational models do not make new predictions
2. Comprehensive cognitive architectures have been developed	2. There is some overfitting, which restricts generalisation to other data sets
3. Computational models are increasingly used to model brain damage	3. It is sometimes hard to falsify computational effects of models
4. Computational cognitive neuroscience is increasingly used to model patterns of brain activity	4. Most computational models de-emphasise motivational and emotional factors
5. The emphasis on parallel processing fits well with functional neuroimaging data	5. Researchers' reluctance to share source codes and models inhibits progress

> **KEY TERM** 🔑
>
> **Meta-analysis**
> A form of statistical
> analysis based on
> combining the findings
> from numerous studies on
> a given research topic.

One answer to problems with replicability is to use **meta-analysis**, in which findings from many studies are combined and integrated using various statistical techniques. This approach has the advantage of not exaggerating the importance of any single study, but has various potential problems (Sharpe, 1997):

(1) The "apples and oranges" problem: very different studies are often included within a meta-analysis.
(2) The "file drawer" problem: it is hard for researchers to publish non-significant findings. Since meta-analyses often ignore unpublished findings, the studies included may be unrepresentative.
(3) The "garbage in – garbage out" problem: poorly designed and conducted studies are often included along with high-quality ones.

The above problems can be addressed. Precise criteria for studies to be included can reduce the first and third problems. The second problem can be reduced by asking researchers to provide relevant unpublished data.

Watt and Kennedy (2017) identified a more important problem: many researchers use somewhat subjective criteria for inclusion of studies in meta-analyses, favouring those supporting their theoretical position and rejecting those that do not. This creates confirmation bias (see Chapter 14).

The good news is that experimental psychologists (including cognitive psychologists) have responded very positively to the above problems. More specifically, there has been a large increase in *disclosure* and *pre-registration*. Disclosure means that researchers "disclose all of their measures, manipulations, and exclusions" (Nelson et al., 2018, p. 518). In the case of meta-analyses, it means researchers making available all the findings initially considered for inclusion. That would allow other researchers to conduct their own meta-analyses and check whether the outcome remains the same.

Pre-registration involves researchers making publicly available all decisions about sample size, hypotheses, statistical analyses and so on *before* an experimental study is carried out. With respect to meta-analyses, pre-registration involves making public the inclusion criteria for a meta-analysis, the methods of analysis and so on, *before* the findings of included studies are known (Hamlin, 2017).

In sum, there are some genuine issues concerning the replicability of findings within cognitive psychology. However, there are various reasons why there is no "replication crisis". First, numerous important findings in cognitive psychology have been replicated dozens or even hundreds of times as is clear from a very large number of meta-analytic reviews. Second, as Nelson et al. (2018, p. 511) pointed out: "The scientific practices of experimental psychologists have improved dramatically."

OUTLINE OF THIS BOOK

One problem with writing a textbook of cognitive psychology is that virtually all the processes and systems in the cognitive system are interdependent. Consider, for example, a student *reading* a book to prepare for an examination. The student is *learning*, but several other processes are going

on as well. *Visual perception* is involved in the intake of information from the printed page, and there is *attention* to the content of the book.

In order for the student to benefit from the book, they must possess considerable *language skill*, and must have extensive relevant knowledge stored in *long-term memory*. There may be an element of *problem solving* in the student's attempts to relate the book's content to the possibly conflicting information they have learned elsewhere. *Decision-making* may also be involved when the student decides how much time to devote to each chapter.

In addition, what the student learns depends on their *emotional state*. Finally, the acid test of whether the student's learning has been effective comes during the examination itself, when the material from the book must be *retrieved* and *consciously* evaluated to decide its relevance to the examination question.

The words italicised in the previous three paragraphs indicate major aspects of human cognition and form the basis of our coverage. In view of the *interdependence* of all aspects of the cognitive system, we emphasise how each process (e.g., perception) depends on other processes and structures (e.g., attention, long-term memory). This should aid the task of understanding the complexities of human cognition.

CHAPTER SUMMARY 📝

- **Introduction.** Cognitive psychology used to be unified by an approach based on an analogy between the mind and the computer. This information-processing approach viewed the mind as a general-purpose, symbol-processing system of limited capacity. Today there are four main approaches to human cognition: experimental cognitive psychology; cognitive neuroscience; cognitive neuropsychology; and computational cognitive science. These four approaches are increasingly combined to provide an enriched understanding of human cognition.

- **Cognitive psychology.** Cognitive psychology focuses on internal mental processes whereas behaviourism focused mostly on observable stimuli and responses. Cognitive psychologists assume top-down and bottom-up processes are both involved in the performance of cognitive tasks. These processes can be serial or parallel. Various methods (e.g., latent-variable analysis) have been used to address the task impurity problem and to identify the processes within cognitive tasks. Cognitive psychology has massively influenced theorising and the tasks used across all major approaches to human cognition. In spite of its enormous contributions, cognitive psychology sometimes lacks ecological validity, suffers from paradigm specificity and is theoretically vague.

- **Cognitive neuropsychology.** Cognitive neuropsychology is based on various assumptions including modularity, anatomical

modularity, uniformity of functional architecture and subtractivity. Double dissociations provide reasonable (limited) evidence for separate modules or systems. The case-study approach is more informative than the single-case approach. Cognitive neuropsychology is limited for several reasons: its assumptions are mostly too strong; patients can develop compensatory strategies, there is brain plasticity; brain damage can cause widespread reduced connectivity within the brain; and it underestimates the extent of integrated brain functioning.

• **Cognitive neuroscience: the brain in action.** Cognitive neuroscientists study the brain as well as behaviour using techniques varying in spatial and temporal resolution. Functional neuroimaging techniques provide basically correlational evidence, but transcranial magnetic stimulation (TMS) can indicate a given brain area is necessarily involved in a given cognitive function. The richness of the data obtained from neuroimaging studies permits the assessment of functional specialisation and brain integration. Cognitive neuroscience is a flexible and potentially self-correcting approach. However, correlational findings are sometimes over-interpreted, underpowered studies make replication difficult, and relatively few studies in cognitive neuroscience generate (or even test) cognitive theories.

• **Computational cognitive science.** Computational cognitive scientists develop computational models to understand human cognition. Connectionist networks use elementary units or nodes connected together. They can learn using rules such as backward propagation. Production systems consist of production or "IF . . . THEN" rules. ACT-R is a highly developed model based on production systems. Computational models have increased in scope to provide detailed theoretical accounts of findings from cognitive neuroscience and cognitive neuropsychology. They have shown progress via the use of nested incremental modelling. Computational models are often hard to falsify, de-emphasise motivational and emotional factors, and often lack biological plausibility.

• **Comparisons of different approaches.** The major approaches are increasingly used in combination. Each approach has its own strengths and limitations, which makes it useful to use converging operations. When two approaches produce the same findings, this is stronger evidence than can be obtained from a single approach on its own. If two approaches produce different findings, this indicates further research is needed to clarify what is happening.

• **Is there a replication crisis?** There is increasing evidence that many findings in psychology (including cognitive psychology) are hard to replicate. However, this does *not* mean there is a

replication crisis. Meta-analyses indicate that numerous findings have been successfully replicated many times. In addition, experimental research practices have improved considerably in recent years which should increase successful replications in the future.

FURTHER READING

Hartwigsen, G. (2018). Flexible redistribution in cognitive networks. *Trends in Cognitive Sciences*, *22*, 687–698. Gesa Hartwigsen discusses several compensatory strategies used by brain-damaged patients and healthy individuals administered transcranial magnetic stimulation (TMS).

Laird, J.E., Lebiere, C. & Rosenbloom, P.S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*, 1–19. These authors proposed a standard model based on the commonalities found among different proposed cognitive architectures (e.g., ACT-R; Soar).

Passingham, R. (2016). *Cognitive Neuroscience: A very Short Introduction*. Oxford: Oxford University Press. Richard Passingham provides an accessible account of the essential features of cognitive neuroscience.

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafo, M.R., et al. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115–126. Problems with research in cognitive neuroscience are discussed and proposals for enhancing the quality and replicability of such research are put forward.

Shallice, T. (2015). Cognitive neuropsychology and its vicissitudes: The fate of Caramazza's axioms. *Cognitive Neuropsychology*, *32*, 385–411. Tim Shallice discusses strengths and limitations of various experimental approaches within cognitive neuropsychology.

Shrout, P.E. & Rodgers, J.L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487–510. Patrick Shrout and Joseph Rodgers discuss the numerous ways in which improving research practices are reducing replication problems.

Taatgen, N.A., van Vugt, M.K., Borst, J.P. & Melhorn, K. (2016). Cognitive modelling at ICCM: State of the art and future directions. *Topics in Cognitive Science*, *8*, 259–263. Niels Taatgen and his colleagues discuss systematic improvements in computational cognitive models.

Ward, J. (2015). *The Student's Guide to Cognitive Neuroscience* (3rd edn). Hove, UK: Psychology Press. The first five chapters of this textbook provide detailed information about the main techniques used by cognitive neuroscientists.

Visual perception and attention

PART I

What is "perception"? According to Twedt and Parfitt (2018, p. 1), "Perception is the study of how sensory information is processed into perceptual experience . . . all senses share the common goal of picking up sensory information from the external environment and processing that information into a perceptual experience."

Our main emphasis in this section of the book is on visual perception, which is of enormous importance in our everyday lives. It allows us to move around freely, to see other people, to read magazines and books, to admire the wonders of nature, to play sports and to watch movies and television. It also helps to ensure our survival. If we misperceive how close cars are to us as we cross the road, the consequences could be fatal. Unsurprisingly, far more of the cortex (especially the occipital lobes at the back of the head) is devoted to vision than to any other sensory modality.

Visual perception seems so simple and effortless, we typically take it for granted. In fact, however, it is very complex, and numerous processes transform and interpret sensory information. Relevant evidence comes from researchers in artificial intelligence who tried to program computers to "perceive" the environment. In spite of their best efforts, no computer can match more than a fraction of the skills of visual perception that we possess. For example, humans are much better than computer programs when deciphering distorted interconnected characters (commonly known as CAPTCHAs) to gain access to an internet website.

There is a rapidly growing literature on visual perception (especially from the cognitive neuroscience perspective). The next three chapters provide detailed coverage of the main issues. Chapter 2 focuses on basic processes in visual perception with an emphasis on the great advances made in understanding the underlying brain mechanisms. Of importance, we see in this chapter that the processes leading to object recognition differ from those guiding vision for action. Finally, this chapter discusses important aspects of visual perception (e.g., colour perception; perception without awareness; depth perception).

Chapter 3 focuses on the processing underlying our ability to identify objects in the world around us. Initially, we discuss perceptual organisation and how we decide which parts of the visual input belong together and so form an object. We then move on to theories of object recognition including a discussion of the relevant behavioural and neuroscientific evidence.

Are the same recognition processes involved across all types of objects? This issue remains controversial. However, most experts agree that face recognition differs in important ways from the recognition of most other objects. Accordingly, face recognition is discussed separately from the recognition of other objects.

The final part of Chapter 3 is concerned with another controversial issue – whether the main processes involved in visual imagery are the same as those involved in visual perception. As you will see, it is arguable that this controversy has largely been resolved (turn to Chapter 3 to find out how!).

The central focus of Chapter 4 is on how we process a constantly changing environment and manage to respond appropriately to those changes. Of major importance is our ability to predict the speed and direction of objects and to move towards our goal whether walking or driving. Other topics discussed in Chapter 4 are our ability to reach for (and grasp) objects and our ability to make sense of other people's movements.

There are major links between visual perception and attention. The final topic in Chapter 4 is concerned with the notion that we may need to attend to an object to perceive it consciously. Attentional failures can prevent us from noticing changes in objects or the presence of an unexpected object. However, failures to notice changes in objects also depend on the limitations of peripheral vision.

Issues relating directly to attention are discussed thoroughly in Chapter 5. This chapter starts with the processes involved in focused attention in the visual and auditory modalities. We next consider how we use visual processes when engaged in the everyday task of searching for some object (e.g., a pair of socks in a drawer). We then consider research on disorders of visual attention in brain-damaged individuals, research that has greatly increased our understanding of visual attention in healthy individuals. After that, we discuss the factors determining the extent to which we can do two things at once (i.e., multi-tasking). This involves a consideration of the role played by "automatic" processes.

In sum, the area spanning visual attention and attention is among the most exciting and important within cognitive psychology and cognitive neuroscience. Tremendous progress has been made in unravelling the complexities of perception and attention over the past decade. The choicest fruits of that endeavour are set before you in the four chapters forming this section of the book.

Basic processes in visual perception

INTRODUCTION

Considerable progress has been made in understanding visual perception in recent years. Much of this progress is due to cognitive neuroscientists, thanks to whom we now have a good knowledge of the visual brain. Initially, we consider the main brain areas involved in vision and their functions. Then we discuss theories of brain systems in vision, followed by a detailed analysis of basic aspects of visual perception (e.g., colour perception; depth perception). Finally, we consider whether perception can occur without conscious awareness.

The specific processes we use in visual perception depend on what we are looking *at* and on our perceptual goals (i.e., what we are looking *for*) (Hegdé, 2008). On the one hand, we can sometimes perceive the gist of a

Figure 2.1
Complex scene that requires prolonged perceptual processing to understand fully. Study the picture and identify the animals within it.

Reprinted from Hegdé (2008). Reprinted with permission of Elsevier.

DOI: 10.4324/9781351058513-3

Interactive feature:
Primal Pictures'
3D atlas of the brain

natural scene extremely rapidly (Thorpe et al., 1996). Observers saw photographs containing (or not containing) an animal for only 20 ms. Event-related potentials (ERPs: see Glossary) indicated the presence of an animal was detected within about 150 ms. On the other hand, look at the photograph in Figure 2.1 and decide how many animals are present. It probably took you several seconds to perform this task. Bear in mind the diversity of visual perception as you read this and the following two chapters.

VISION AND THE BRAIN

In this section we consider the brain systems involved in visual perception. Visual processing occurs in at least 30 distinct brain areas (Felleman & Van Essen, 1991). The visual cortex consists of the entire occipital cortex at the back of the brain and also extends well into the temporal and parietal lobes. To understand visual processing in the brain fully, however, we first need to consider briefly what happens between the eye and the cortex.

From eye to cortex

There are two types of visual receptor cells in the retina: cones and rods. Cones are used for colour vision and sharpness of vision (see section on colour vision, pp. 64–71). Patients with rod monochromatism have no detectable cone function resulting in total colour blindness (Tsano & Sharma, 2018).

There are 125 million rods concentrated in the outer regions of the retina. Rods are specialised for vision in dim light. Many differences between cones and rods stem from the fact that a retinal ganglion cell receives input from only a few cones but from hundreds of rods. Thus, only rods produce much activity in retinal ganglion cells in poor lighting conditions.

The main pathway between the eye and the cortex is the retina-geniculate-striate pathway. It transmits information from the retina to V1 and then V2 (both discussed shortly) via the lateral geniculate nuclei (LGNs) of the thalamus. The entire retina-geniculate-striate system is organised similarly to the retinal system. For example, two stimuli adjacent to each other in the retinal image will also be adjacent at higher levels within that system. The technical term is retinopy: retinal receptor cells are mapped to points on the surface of the visual cortex.

Each eye has its own optic nerve and the two optic nerves meet at the optic chiasm. At this point the axons from the outer halves of each retina proceed to the hemisphere on the brain hemisphere on the same side, whereas those from the inner halves cross over and go to the other hemisphere. As a result, each side of visual space is represented within the opposite brain hemisphere. Signals then proceed along two optic tracts within the brain. One tract contains signals from the left half of each eye and the other signals from the right half (see Figure 2.2).

After the optic chiasm, the optic radiation proceeds to the lateral geniculate nucleus, which is part of the thalamus. Nerve impulses finally reach V1 in primary visual cortex within the occipital lobe at the back of the head before spreading out to nearby visual cortical areas such as V2.
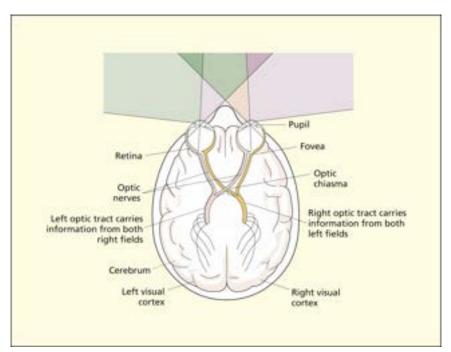
There are two relatively independent channels or pathways within the retina-geniculate-striate system:

(1) The parvocellular (or P) pathway: it is most sensitive to colour and to fine detail; most of its input comes from cones.
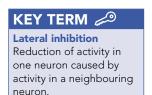(2) The magnocellular (or M) pathway: it is most sensitive to movement information; most of its input comes from rods.

As stated above, these two pathways are only relatively independent. In fact, there are numerous interconnections between them and the entire visual system is extremely complex. For example, there is clear intermingling of the two pathways in V1 (Leopold, 2012). Ryu et al. (2018, p. 707) studied brain activity in V1 when random-dot images were presented: "The local V1 sites receiving those parallel inputs [from the P and M pathways] are densely linked with one another via horizontal connections [which] are organised in complicated yet systematic ways to subserve the multitude of representational functions of V1."

Finally, there is also a Koniocellular pathway. However, its functions are still not well understood.

Early visual processing: V1 and V2

We start with three general points. First, to understand visual processing in the primary visual cortex (V1 or BA17) and the secondary visual cortex (V2 or BA18), we must consider the notion of a receptive field. The receptive field for any given neuron is the retinal region where light affects its activity. The receptive field can also refer to visual space because it is mapped in a one-to-one manner onto the retinal surface.

KEY TERM

Receptive field
The region of the retina in which light influences the activity of a particular neuron.

KEY TERM 🔑

Lateral inhibition
Reduction of activity in one neuron caused by activity in a neighbouring neuron.

Second, neurons often influence each other. For example, there is lateral inhibition, where reduced activity in one neuron is caused by activity in a neighbouring neuron. Lateral inhibition increases the contrast at the edges of objects, making it easier to identify the dividing line between objects. The phenomenon of simultaneous contrast depends on lateral inhibition (see Figure 2.3). The two central squares are physically identical but the one on the left appears lighter. This difference is due to simultaneous contrast produced because the left surround is much darker than the right surround.

Third, early visual processing involves large areas within the primary visual cortex (V1) and secondary visual cortex (V2). For example, Hegdé and Van Essen (2000) found in macaques that one-third of V2 cells responded to complex shapes and differences in size and orientation.

Figure 2.3
The square on the right looks darker than the identical square on the left because of simultaneous contrast involving lateral inhibition.

From Lehar (2008). Reproduced with permission of the author.

Two pathways

As we have just seen, neurons from the P and M pathways mainly project to V1 (primary visual cortex). What happens after V1? The P pathway associates with the ventral pathway or stream that proceeds to the inferotemporal cortex. In contrast, the M pathway associates with the dorsal pathway or stream that proceeds to the posterior parietal cortex. Note the above assertions oversimplify a complex reality.

We discuss the ventral and dorsal pathways in detail shortly. It is assumed the ventral or "what" pathway culminating in the inferotemporal cortex is mainly concerned with form and colour processing and with object recognition. In contrast, the dorsal or "how" pathway culminating in the parietal cortex is more concerned with motion processing. As we will see later, there are extensive interactions between the two pathways. The nature of such interactions was reviewed by Rossetti et al. (2017; see Figure 2.15 in this chapter).

Galletti and Fattori (2018) argued that visual processing is more *flexible* than implied by the notion of two interacting pathways or streams.

> We should not conceive the cortical streams as fixed series of interconnected cortical areas in which each area belongs to one stream . . ., but [rather] as interconnected neuronal networks, often involving the same neurons, that are involved in a number of functional processes and whose activation changes dynamically according to the context.
>
> (p. 203)

Organisation of the visual brain

A more detailed picture of the brain areas involved in visual processing is given in Figure 2.4. V3 is generally assumed to be involved in form
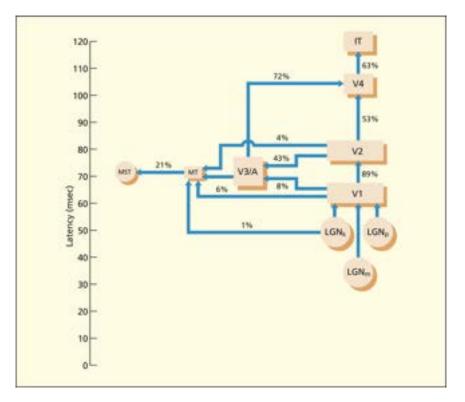
Figure 2.4
Some distinctive features of the largest visual cortical areas. The relative size of the
boxes reflects the relative area of different regions. The arrows labelled with percentages
show the proportion of fibres in each projection pathway. The vertical position of each
box represents the response latency of cells in each area, as measured in single-unit
recording studies. IT = inferotemporal cortex; MT = medial or middle temporal cortex;
MST = medial superior temporal cortex. All areas are discussed in detail in the text.
From Mather (2009). Copyright 2009 George Mather. Reproduced with permission.

processing, V4 in colour processing and V5/MT in motion processing (all
discussed in more detail pp. 49–54). The ventral stream includes V1, V2, V3,
V4 and the inferotemporal cortex, whereas the dorsal stream proceeds from
V1 via V3 and MT (medial temporal cortex) to MST (medial superior tem-
poral cortex).

Figure 2.4 reveals three important points. First, there are complex
interconnections among visual cortical areas. Second, the brain areas
within the ventral pathway are more than twice as large as those within
the dorsal pathway. Third, cells in the lateral geniculate nucleus respond
fastest when a visual stimulus is presented followed by activation of cells
in V1. However, cells are activated in several other areas (V3/V3A; MT;
MST) very shortly thereafter.

Figure 2.4 shows the traditional *hierarchical* view of the major brain
areas involved in visual processing. It is supported by anatomical evi-
dence (see proportions of fibres projecting up the hierarchy in the figure).
Nevertheless, this view is oversimplified. Here we consider three of its main
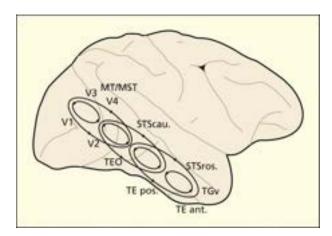limitations.

Figure 2.5

Connectivity within the ventral pathway on the lateral surface of the macaque brain. Brain areas involved include V1, V2, V3 and V4, the middle temporal (MT)/medial superior temporal (MST) complex, the superior temporal sulcus (STS) and the inferior temporal cortex (TE).

From Kravitz et al. (2013). Reprinted with permission of Elsevier.

First, Kravitz et al. (2013) disagreed with the traditional view that the ventral pathway or stream involves a serial hierarchy proceeding from simple to complex. Instead, he argued it consists of several overlapping recurrent *networks* (see Figure 2.5). There are connections in both directions between the components within these networks. As Hegdé (2018, p. 902) argued, "Various regions of the visual system process information not in a strict hierarchical manner but as parts of various dynamic brain-wide networks."

Second, there is an initial "feedforward sweep" proceeding through the visual areas starting with V1 and then V2 (shown by the directional arrows in Figure 2.4). This is followed by recurrent or top-down processing proceeding in the opposite direction (not shown in Figure 2.4). Several theorists (e.g., Lamme, 2018; see Chapter 16) assume recurrent processing is of major importance for conscious visual perception because it *integrates* information across different visual areas. Note that visual imagery depends on several top-down processes resembling those used in visual perception (see Chapter 3).

Hurme et al. (2017) obtained support for the above assumptions. They applied transcranial magnetic stimulation (TMS; see Glossary) to V1 at 60 ms to suppress feedforward processing and at 90 ms to suppress recurrent processing. As predicted, *early* V1 activity was necessary for conscious and unconscious vision but *late* V1 activity was necessary only for conscious vision.

Third, Zeki (2016) distinguished three hierarchical models of the visual brain (see Figure 2.6). Model (a) was proposed first and model (c), the one favoured by Zeki, was proposed most recently. His central argument is that, "Parallel processing . . . is much more ubiquitous than commonly supposed" (p. 2515). Thus, models such as the one shown in Figure 2.4 are inadequate because they de-emphasise parallel processing.

Functional specialisation

Zeki (1993, 2001) proposed a functional specialisation theory where different cortical areas are specialised for different visual functions. The visual system resembles workers each working alone to solve part of a complex problem, and it is consistent with Zeki's (2016) emphasis on parallel processing within the visual brain. The results of their labours are then combined to produce coherent visual perception.

What are the advantages of functional specialisation? First, object attributes can occur in unpredictable combinations (Zeki, 2005). For example, a green object may be a car, a sheet of paper or a leaf, and a car can be red, black or green. Thus, we often need to process *all* of an
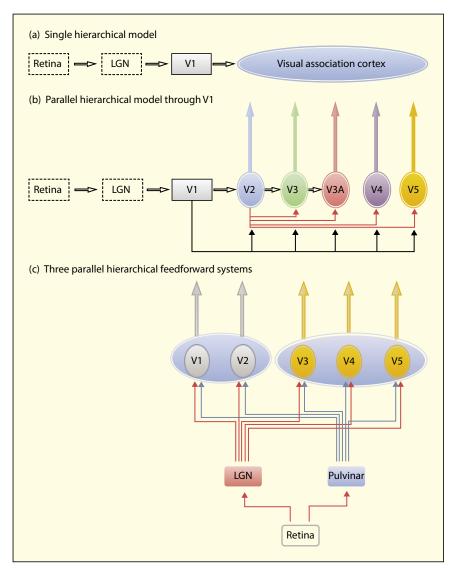
Figure 2.6

(a) The single hierarchical model where all brain areas after V1 are considered jointly as "visual association cortex"; (b) the parallel hierarchical model which is a hierarchy of processing areas running serially from V1 through V2 to V3 but with much parallel processing; (c) the three parallel hierarchical feedforward systems model with a strong emphasis on parallel rather than serial processing.

From Zeki (2016).

object's attributes for accurate perception. Second, the required processing differs considerably across attributes (Zeki, 2005). For example, motion processing involves integrating information across *time* whereas form or shape processing involves considering the *spatial* relationship of elements at a given moment.

Here are the main functions Zeki ascribed to the brain areas shown in Figure 2.3:

- V1 and V2: They are involved at an early stage of visual processing. They contain different groups of cells responsive to colour and form.
- V3 and V3A: Cells in these areas respond to form (especially the shapes of objects in motion) but not colour.
- V4: The majority of cells in this area respond to colour; many are also responsive to line orientation.

- V5: This area is specialised for visual motion. In studies with macaque monkeys, Zeki found all the cells in this area responded to motion but not colour. In humans, the areas specialised for visual motion are referred to as MT and MST.

Zeki assumed colour, motion and form are processed in anatomically separate visual areas. The relevant evidence is discussed below.

Form processing

Brain areas involved in form processing in humans include V1, V2, V3 and V4, culminating in the inferotemporal cortex (Kourtzi & Connor, 2011). Neurons in the inferotemporal cortex respond to specific semantic categories (e.g., animals; body parts; see Chapter 3). Neurons in the inferotemporal cortex are also involved in form processing. Baldassi et al. (2013) found in monkeys that many neurons within the anterior inferotemporal cortex responded on the basis of form or shape (e.g., round; star-like; horizontally thin) rather than object category.

Pavan et al. (2017) investigated the role of early visual areas in form processing using repetitive transcranial magnetic stimulation (rTMS; see Glossary) to disrupt processing. With static stimuli, rTMS delivered to early visual areas (V1/V2) disrupted form processing whereas rTMS delivered to V5/MT did not.

If form processing occurs in different brain areas from colour and motion processing, we might anticipate some patients would have severely impaired form processing but intact colour and motion processing. Some support was reported by Gilaie-Dotan (2016a). She studied LG, a man with visual form agnosia (see Glossary). LG had deficient functioning within V2 and V3 (although no obvious brain damage) associated with impaired form processing and object recognition, but relatively intact perception of colour and biological motion. However, such cases are very rare. As Zeki (1993) pointed out, brain damage sufficient to almost eliminate form perception would typically be so widespread that the patient would be blind.

Colour processing

The assumption that V4 (located within the ventral visual pathway) is specialised for colour processing has been tested in several ways. These include studying brain-damaged patients, using brain-imaging techniques, and using transcranial magnetic stimulation to produce a temporary "lesion" (see pp. 20–22).

If V4 is specialised for colour processing, patients with damage to that area should exhibit minimal colour perception with fairly intact form and motion perception and ability to see fine detail. This is approximately the case in achromatopsia (also known as cerebral achromatopsia) although cases involving total achromatopsia are very rare (Zihl & Heywood, 2016).

Bouvier and Engel (2006) found in a meta-analysis that a small brain area within the ventral (bottom) occipital cortex in (or close to) area V4 was damaged in nearly all cases of achromatopsia. However, the loss of

colour vision was typically only partial, implying other areas are also directly involved in colour processing.

Lafer-Sousa et al. (2016) identified three brain areas in the ventral visual pathway responding more strongly to video clips of various objects presented in colour than in black-and-white. Of importance, different brain areas were associated with colour and shape processing: colour areas responded comparably to intact and scrambled objects.

Bannert and Bartels (2018) studied brain activity in several brain areas (including V1, V2, V3 and V4) while participants viewed abstract colour stimuli or formed visual images of colour objects (e.g., *tomato*; *banana*). The colour of visually presented stimuli could be worked out from brain activity in every brain area studied, whereas the colour of imagined stimuli could only be worked out from brain activity in V4. These findings suggest that a network including several brain areas is involved in colour processing, but V4 is of special importance within that network.

Finally, note that V4 is a relatively large area. As such, it is involved in the processing of texture, form and surfaces as well as colour (Winawer & Witthoft, 2015).

Motion processing

Area V5 (also known as motion processing area MT) is heavily involved in motion processing. Functional neuroimaging studies indicate that motion processing is associated with activity in V5/MT (Zeki, 2015). However, such studies cannot show V5 (or MT) is *necessary* for motion perception. More direct evidence was reported by McKeefry et al. (2008) using transcranial magnetic stimulation (see Glossary) to disrupt motion perception. TMS, applied to V5/MT, produced a subjective slowing of stimulus speed and impaired observers' ability to discriminate between different speeds.

Further evidence of the causal role of V5 in motion perception was obtained by Vetter et al. (2015). Observers could not predict the motion of a moving target when TMS was applied to V5.

Additional evidence that the area V5/MT is important in motion processing comes from research on patients with akinetopsia. **Akinetopsia** is an exceptionally rare condition where stationary objects are perceived fairly normally but motion perception is grossly deficient (Ardila, 2016). Zihl et al. (1983) studied LM, a woman with akinetopsia who had suffered bilateral damage to the motion area (V5/MT). She could locate stationary objects by sight, had good colour discrimination and her binocular vision was normal. However, her motion perception was grossly deficient:

> She had difficulty . . . in pouring tea or coffee into a cup because the fluid appeared to be frozen, like a glacier. . . . In a room where more than two people were walking, . . . "people were suddenly here or there but I have not seen them moving".

Zihl and Heywood (2015) discussed additional findings relating to LM. Even though her motion perception was extremely poor, she still retained limited ability to distinguish between moving and stationary stimuli.

Heutink et al. (2018) studied TD, a patient with akinestopsia due to damage to V5. She was severely impaired at perceiving the direction of

high-speed visual motion but not low-speed motion, suggesting V5 is less important for processing low-speed than high-speed motion.

V5 (MT) is not the only brain area involved in motion processing. There is also the area MST just above V5/MT. Vaina (1998) studied two patients with damage to MST. Both patients had various problems relating to motion perception. One patient (RR) "frequently bumped into people, corners and things in his way, particularly into moving targets (e.g., people walking)" (Vaina, 1998, p. 498). These findings suggest MST is involved in the visual guidance of walking. Chaplin et al. (2018) found that the direction of motion of a stimulus in healthy individuals could be inferred by taking account of activation within both MT and MST.

The notions that motion perception depends almost exclusively on V5/MT and MST and that those areas only process information relating to motion are both oversimplifications for various reasons. First, several areas outside V5/MT and MST are involved in motion perception. For example, consider biological motion perception (see Chapter 4). Such perception involves several additional areas including the superior temporal sulcus, superior temporal gyrus and inferior frontal gyrus (Thompson & Parasuraman, 2012; Pavlova et al., 2017).

Second, Heywood and Cowey (1999; see Figure 2.7) found that approximately 60% of cells within V5/MT respond to binocular disparity (difference between the retinal images in the left and right eyes; see Glossary) and 50% of cells within V5/MT respond to stimulus orientation. However, V5/MT is especially important with respect to direction of motion with approximately 90% of cells responding.

Third, we should distinguish between different types of motion perception (e.g., first-order and second-order motion perception). With first-order
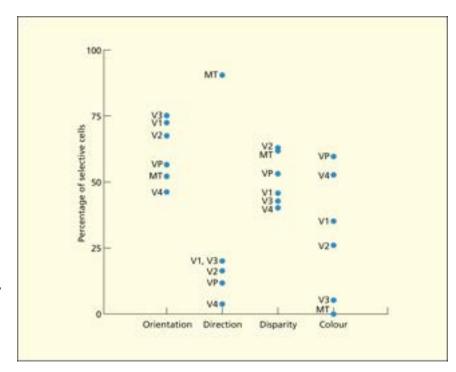

Figure 2.7

The percentage of cells in six different visual cortical areas responding selectively to orientation, direction of motion, disparity and colour.
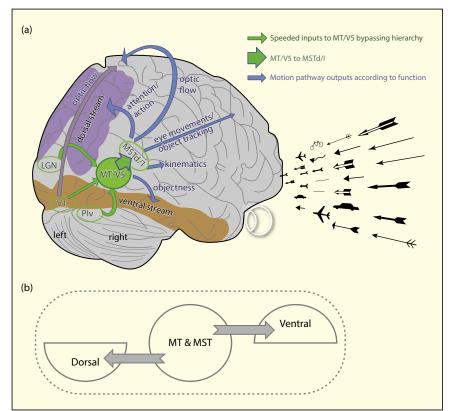
From Heywood and Cowey (1999).

displays, the moving shape differs in luminance (intensity of reflected light) from its background. For example, the shape might be dark whereas the background is light. With second-order displays, there is no difference in luminance between the moving shape and the background.

In everyday life, we encounter second-order displays infrequently (e.g., movement of grass in a field caused by the wind). Some patients have intact first-order motion perception but impaired second-order motion perception whereas others exhibit the opposite pattern (Gilaie-Dotan, 2016a). Thus, all forms of motion perception do not involve similar underlying processes.

Gilaie-Dotan (2016b) studied motion processing within the brain (see Figure 2.8). In essence, information about visual motion inputs bypasses several visual areas (e.g., V2, V4) and rapidly reaches MT/V5. It then continues rapidly to MST, after which information is transferred to several other areas.

Gilaie-Dotan (2016b, p. 379) pointed out that visual motion perception "has been continually associated with and considered part of the dorsal pathway". For example, Milner and Goodale's (1995, 2008) perception-action model emphasises close links between visual motion processing and perception and the dorsal ("how" pathway; see pp. 56–57). Gilaie-Dotan accepted the dorsal pathway is of major importance. However, she argued persuasively that the ventral ("what") pathway is also involved in motion perception. For example, efficient detection of visual motion

requires having an intact right ventral visual cortex (Gilaie-Dotan et al., 2013).

Binding problem

Zeki's theoretical approach poses the obvious problem of how information about an object's motion, colour and form is combined and integrated to produce coherent perception. This is the **binding problem**: "How the brain brings together what it has processed . . . in its different hierarchically organised parallel processing systems . . . to give us our unitary experience of the visual world" (Zeki, 2016, p. 3521). One aspect of this problem is that object-related processing in different visual areas ends at different times, thus making it harder to integrate these outputs in visual perception.

There may be continuous integration of information starting during early stages of visual processing. Seymour et al. (2009) presented observers with red or green dots rotating clockwise or counterclockwise. Colour-motion conjunctions were processed in several brain areas including V1, V2, V3, V3A/B, V4 and V5/MT+. Seymour et al. (2016) found that binding of information about object form and colour occurred as early as V2. These findings contradict the traditional assumption that "The visual system initially extracts borders between objects and their background and then 'fills in' colour" (Seymour et al., 2016, p. 1997).

Ghose and Ts'o (2017) reviewed research indicating progressively more integration of different kinds of information (and thus less functional specialisation) during visual processing. They concluded:

> In V2, we see an increase in the overlap of cortical generated selectivities such as orientation and colour . . . in V4 we see extensive overlap among colour, size, and form, and the existence of . . . a combination of colour and orientation . . . not present in earlier areas.
>
> (p. 17)

So far we have focused on increases in integration of information as processing proceeds from early to late visual areas (*feedforward* processing). However, conscious visual perception generally depends crucially on *recurrent* processing (feedback from higher to lower visual brain areas). Observers' expectations influence recurrent processing and it is arguable that expectations (e.g., bananas will be yellow) facilitate the binding or integration of different kinds of visual information.

The binding-by-synchrony hypothesis (e.g., Singer & Gray, 1995) provides an influential solution to the binding problem. According to this hypothesis, detectors responding to features of a single object fire in synchrony. Of relevance, widespread synchronisation of neural activity is associated with conscious visual awareness (e.g., Gaillard et al., 2009; Melloni et al., 2007; see Chapter 16).

The synchrony hypothesis is oversimplified. There is the largely unresolved issue of explaining *why* and *how* synchronised activity occurs across visual areas. The fact that visual object processing occurs in widely distributed areas of the brain makes it implausible that precise synchrony could be achieved.

Finally, note there are various binding problems. As Feldman (2013) pointed out, one problem is how visual features are bound together. Another problem is how we bind together information over successive eye movements to perceive a stable visual world.

Within the above broader context, it is clear several lines of research are relevant. For example, observers must decide which parts of the visual information available at any given time belong to the same object. The gestaltists put forward several laws describing how this happens (see Chapter 3). Research on visual search (detecting target stimuli among distractors) is also relevant (see Chapter 5). This research shows the important role of selective attention in combining features close together in time and space.

<div style="float:right; border:1px solid #888; padding:6px; width:200px;">

KEY TERM 🔑

Ventral stream
The part of the visual processing system involved in object perception and recognition and the formation of perceptual representations.

</div>

Evaluation

Zeki's functional specialisation theory is an ambitious and influential attempt to provide a coherent theoretical framework. As discussed later, Zeki's assumption that motion processing typically proceeds somewhat independently of other types of visual process has reasonable empirical support.

What are the limitations with Zeki's theoretical approach? First, the brain areas involved in visual processing are less specialised than implied theoretically. As mentioned earlier on p. 52, Heywood and Cowey (1999) considered the percentage of cells in each visual cortical area responding selectively to various stimulus characteristics (see Figure 2.7). Cells in several areas responded to orientation, disparity and colour. Specialisation was found only with respect to responsiveness to direction of stimulus motion in MT.

Second, the visual brain is substantially more complex than assumed by Zeki. There are far more brain areas devoted to visual processing than shown in Figure 2.3, and each brain area has connections to numerous other areas (Baker et al., 2018). For example, V1 is connected to at least 50 other areas! What is also de-emphasised in Zeki's approach is the importance of brain networks and the key role played by recurrent processing.

Third, the binding problem (or problems) has not been solved. However, integrated visual perception undoubtedly depends on both bottom-up (feedforward) processes and top-down (recurrent) processes (see Chapter 3).

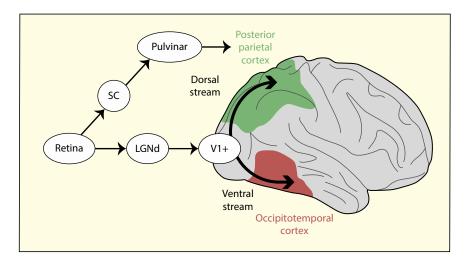## TWO VISUAL SYSTEMS: PERCEPTION-ACTION MODEL

What are the major functions of the visual system? Historically, the most popular answer was that it provides us with an internal (and conscious) representation of the external world. In contrast, Goodale and Milner (1992) and Milner and Goodale (1995, 2008) argued in their perception-action model, there are two visual systems each fulfilling a different function or purpose. First, there is the vision-for-perception (or "what") system based on the **ventral stream** or pathway. It is used when we decide whether an object is a cat or a buffalo or when admiring a magnificent landscape. Thus, it is used to identify objects.

Figure 2.9

Goodale and Milner's (1992) perception-action model showing the dorsal and ventral streams. (SC = superior colliculus; LGNd = dorsal lateral geniculate nucleus; V1+ = early visual areas.)

From de Haan et al. (2018). Reprinted with permission of Elsevier.

Second, there is the vision-for-action (or "how") system based on the **dorsal stream** or pathway (see Figure 2.9) used for visually guided action. It is used when running to return a ball at tennis or when grasping an object. When we grasp an object, we must calculate its orientation and position with respect to ourselves. Since observers and objects often move relative to each other, orientation and position need to be worked out immediately prior to initiating a movement.

Milner (2017, p. 1297) summarised key differences between the two systems:

> The ventral stream . . . mediates the transformations of the contents of the visual signal into the mental furniture that guides memory, recognition and conscious perception. In contrast, the dorsal stream . . . mediates the visual guidance of action, primarily in real time.

Schenk and McIntosh (2010) identified four major differences between the two processing streams:

(1) The ventral stream underlies vision for perception whereas the dorsal stream underlies vision for action.
(2) There is **allocentric coding** (object-centred; coding the locations of objects relative to each other) in the ventral stream but **egocentric coding** (body-centred; coding relative to the observer's own body) in the dorsal stream.
(3) Representations in the ventral stream are sustained over time whereas those in the dorsal stream are short-lasting.
(4) Processing in the ventral stream generally (but not always) leads to conscious awareness, whereas processing in the dorsal stream does not.

Two other differences have been suggested. First, processing in the dorsal stream is *faster*. Second, ventral stream processing depends more on input from the fovea (the central part of the retina used for detecting detail).

Milner and Goodale originally implied that the dorsal and ventral streams were largely independent of each other. However, they have

KEY TERMS 🔑

Dorsal stream
The part of the visual processing system most involved in visually guided action.

Allocentric coding
Visual or spatial coding of objects relative to each other; see **egocentric coding**.

Egocentric coding
Visual or spatial coding dependent on the position of the observer's body; see **allocentric coding**.

increasingly accepted that the two streams often interact. For example, Milner argued the influence of the ventral stream on dorsal stream processing "seems to carry visual and semantic complexity, thereby allowing us to bring meaning to our actions" (Milner, 2017, p. 1305). The key issue of the independence (or interdependence) of the two streams is discussed on pp. 62–63.

Findings: brain-damaged patients

We can test Milner and Goodale's theory by studying brain-damaged patients. Patients with damage to the dorsal pathway should have reasonably intact vision for perception but severely impaired vision for action. The opposite pattern of intact vision for action but very poor vision for perception should be found in patients having damage to the ventral pathway. Thus, there should be a double dissociation (see Glossary).

Optic ataxia

Patients with optic ataxia have damage to the posterior parietal cortex (forming part of the dorsal stream; see Figure 2.10). Some evidence suggests patients with **optic ataxia** are poor at making precise visually guided movements although their vision and ability to move their arms are reasonably intact. As predicted, Perenin and Vighetto (1988) found patients with optic ataxia had great difficulty in rotating their hands appropriately when reaching towards (and into) a large oriented slot.

Patients with optic ataxia do not all conform to the simple picture described above. First, somewhat different regions of posterior parietal cortex are associated with reaching and grasping movements and some patients have greater problems with one type of movement than the other (Vesia & Crawford, 2012).

Second, it is oversimplified to assume patients have intact visual perception but impaired visually guided action. Pisella et al. (2006) obtained much less evidence for impaired visually guided action in central compared to peripheral vision. This finding is consistent with evidence indicating many optic ataxics can drive effectively.
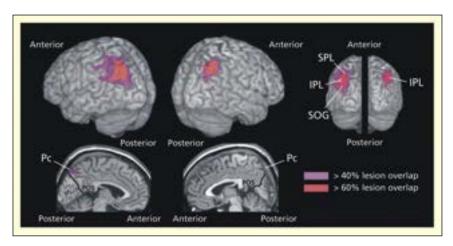
Figure 2.10
Lesion overlap (purple = >40% overlap; orange = >60% overlap) in patients with optic ataxia. (SPL = superior parietal lobule; SOG = superior occipital gyrus; Pc = precuneus.)

From Vesia and Crawford (2012). Reprinted with permission of Springer.

Third, patients with optic ataxia have some impairment in vision for perception (especially in peripheral vision). Bartolo et al. (2018) found such patients had an impaired ability on the perceptual task of deciding whether a target was reachable and they also had problems on tasks requiring vision for action. Thus, patients with optic ataxia have difficulties in combining information from the dorsal and ventral streams.

Fourth, Rossetti and Pisella (2018) concluded as follows from their review: "Optic ataxia is not a visuo-motor deficit and there is no dissociation between perception and action capacities in optic ataxia" (p. 225).

Visual form agnosia

What about patients with damage only to the ventral stream? Of relevance are some patients with **visual form agnosia**, a condition involving severe problems with object recognition even though visual information reaches the visual cortex (see Chapter 3). The most-studied visual form agnosic is DF, whose brain damage is in the ventral stream (James et al., 2003). For example, her activation in that stream was no greater when presented with object drawings than with scrambled line drawings. However, she showed high levels of activation in the dorsal stream when grasping for objects.

Goodale et al. (1994) found DF was very poor at a visual perception task that involved distinguishing between two shapes with irregular contours. However, she grasped these shapes firmly between her thumb and index finger. Goodale et al. concluded DF "had no difficulty in placing her fingers on appropriate opposition points during grasping" (p. 604).

Himmelbach et al. (2012) re-analysed DF's performance based on data in Goodale et al. (1994). DF's performance was substantially inferior to that of healthy controls. Similar findings were obtained when DF's performance on other grasping and reaching tasks was compared against controls. Thus, DF had greater difficulties with visually guided action than previously believed.

Rossit et al. (2018) found DF had impaired peripheral (but not central) reaching, which is the pattern associated with optic ataxia. DF also had significant impairment in the fast control of reaching movements (also associated with optic ataxia). Rossit et al. (p. 15) concluded: "We can no longer assume that DF's dorsal visual stream is intact and that she is spared in visuo-motor control tasks, as she also presents clear signs of optic ataxia."

Visual illusions

There are numerous visual illusions, of which the Müller-Lyer (see Figure 2.11) is one of the most famous. The vertical line on the left looks longer than the one on the right although they are the same length. The Ebbinghaus illusion (see Figure 2.12) is also well known. The central circle surrounded by smaller circles looks smaller than a central circle of the same size surrounded by larger circles although the two central circles are the same size.
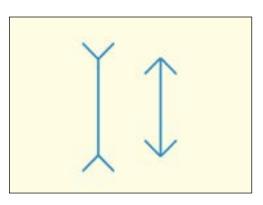
Interactive exercise:
Müller-Lyer

Figure 2.11
The Müller-Lyer illusion.

How has the human species flourished if our visual perceptual processes are apparently very prone to error? Milner and Goodale (1995) argued the vision-for-perception system processes visual illusions and provides visual judgements. In contrast, we mostly use the vision-for-action system when walking close to a precipice or dodging cars. These ideas led to a dramatic prediction: actions (e.g., pointing; grasping) using the vision-for-action system should be unaffected by most visual illusions.

Findings

Bruno et al. (2008) conducted a meta-analytic review of Müller-Lyer studies where observers *pointed* rapidly at one figure (using the vision-for-action system). The mean illusion effect was 5.5%. In contrast, the mean illusion effect was 22.4% when observers provided verbal estimations of length (using the vision-for-perception system). The perception-action model is supported by this large difference. However, the model seems to predict there should have been no illusion effect at all with pointing.

With the Ebbinghaus illusion, the illusion is often much stronger with visual judgements
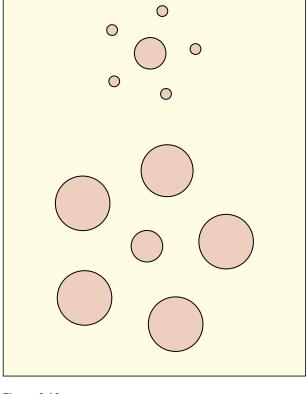

Figure 2.12
The Ebbinghaus illusion.

using the vision-for-perception system than with grasping movements using the vision-for-action system (Whitwell & Goodale, 2017). Knol et al. (2017) explored the Ebbinghaus illusion in more detail. As predicted theoretically, only visual judgements were influenced by the *distance* between the target and the context.

Support for the perception-action model has been reported with the hollow-face illusion, a realistic hollow mask resembling a normal face (see Figure 2.13; visit the website: www.richardgregory.org/experiments). Króliczak et al. (2006) placed a target (a small magnet) on the face mask or a normal face. Here are two tasks they used:

(1) Draw the target position (using the vision-for-perception system).
(2) Make a fast, flicking finger movement to the target (using the vision-for-action system).
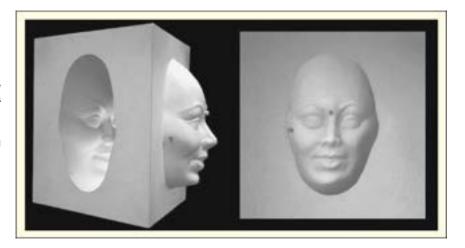
There was a strong illusion effect when observers drew the target position, whereas their performance was very accurate (i.e., illusion-free) when they made a flicking movement. Both findings were as predicted theoretically.

Króliczak et al. (2006) also had a third condition where observers made a *slow* pointing finger movement to the target and so the vision-for-action system was involved. However, there was a fairly strong illusory effect. Why was this? Actions may involve the vision-for-perception system

> **KEY TERM**
>
> **Hollow-face illusion**
> A concave face mask is misperceived as a normal face when viewed from several feet away.

Figure 2.13
Left: normal and hollow faces with small target magnets on the forehead and cheek of the normal face. Right: front view of the hollow mask that appears as an illusory face projecting forwards.

Króliczak et al. (2006). Reprinted with permission of Elsevier.

as well as the vision-for-action system when preceded by conscious cognitive processes.

Various problematical issues for the perception-action model have accumulated. First, the type of action is important. Franz and Gegenfurtner (2008) found the mean illusory effect with the Müller-Lyer was 11.2% with perceptual tasks, compared to 4.4% with full visual guidance of the hand movement. In contrast, grasping when observers could not monitor their hand movements was associated with an illusory effect of 9.4%, perhaps because action programming required the ventral stream.

Second, illusion effects assessed by grasping movements often decrease with repeated practice (Kopiske et al., 2017). Kopiske et al. argued people use *feedback* from their inaccurate grasping movements on early trials to reduce illusion effects later on.

Third, illusion effects are often greater when grasping or pointing movements are made following a delay (Hesse et al., 2016). The ventral stream (vision-for-perception) may be more likely to be involved after a delay.

The various interpretive problems with previous research led Chen et al. (2018a) to use a different approach. In their key condition, observers had restricted vision (they viewed a sphere coated in luminescent paint in darkness through a pinhole). They estimated the sphere's size by matching the distance between their thumb and forefinger to that size (perception) or they grasped the sphere (action). Their non-grasping hand was in their lap or directly below the sphere. In the latter condition, observers could make use of **proprioception** (awareness of the position of one's body parts).

Size judgements were very accurate in perception and action with full vision (see Figure 2.14). However, the key finding was that proprioceptive information about distance produced almost perfect performance when observers grasped the sphere but *not* when providing a perceptual estimate. These findings indicate a very clear difference in the processes underlying vision-for-perception and vision-for-action.

In sum, there is some support for the predictions of the original vision-action model. However, illusory effects with visual judgements and with actions are more complex and depend on many more factors than

KEY TERM 🗝

Proprioception
An individual's awareness of the position and orientation of parts of their body.

assumed by that model. Attempts by Milner and Goodale to accommodate such complexities are discussed below.

Action planning + motor responses

Milner and Goodale (2008) argued most tasks requiring observers to grasp an object involve some processing in the ventral stream in addition to the dorsal stream. They reviewed research showing that involvement of the ventral stream is especially likely in the following circumstances:
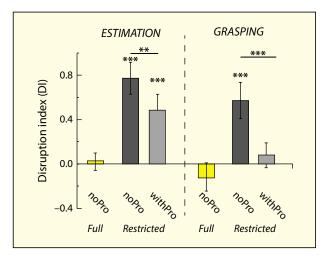

Figure 2.14

Disruption of size judgements when estimated perceptually (estimation) or produced by grasping (grasping) in full or restricted vision when there was proprioception (withPro) or no proprioception (noPro).

From Chen et al. (2018a). Reprinted with permission of Elsevier.

(1) Memory is required (e.g., there is a time lag between the offset of the stimulus and the start of the grasping movement).
(2) Time is available to plan the forthcoming movement (e.g., Króliczak et al., 2006).
(3) Planning which movement to make is necessary.
(4) The action is unpractised or awkward.

According to the perception-action model, actions are most likely to require the ventral stream when they involve conscious processes. Creem and Proffitt (2001) supported this notion. They started by distinguishing between *effective* and *appropriate* grasping. For example, we can grasp a toothbrush effectively by its bristles but appropriate grasping involves accessing stored knowledge about the object and so often requires the ventral stream. As predicted, appropriate grasping was much more adversely affected than effective grasping by disrupting participants' ability to retrieve object knowledge.

van Polanen and Davare (2015) reviewed research on factors controlling skilled grasping. They concluded:

> The ventral stream seems to be gradually more recruited as information about the object from pictorial cues or memory is needed to control the grasping movement, or if conceptual knowledge about more complex objects that are used every day or tools needs to be retrieved for allowing the most appropriate grasp.
>
> (p. 188)

Dorsal stream: conscious awareness

According to the two systems approach, ventral stream processing is generally accessible to consciousness whereas dorsal stream processing is not. For example, it is assumed that the ventral stream (and conscious processing) are often involved in motor *planning* (Milner & Goodale, 2008). There is some support for these predictions from the model (Milner, 2012). As we will see, however, recent evidence mostly provides contrary evidence.

Ludwig et al. (2016) assessed the involvement of the dorsal and ventral streams in conscious visual perception using a different approach. The visibility of visual targets presented to one eye was manipulated by varying the extent to which continuous flash suppression (rapidly changing stimuli presented to the other eye) impaired the processing of the targets.

There were two main findings. First, there was a tight coupling between visual awareness of target stimuli and ventral stream processing. Second, there was a much looser coupling between target awareness and dorsal stream processing. The first finding is consistent with the two visual systems hypothesis. However, the second finding suggests dorsal processing is more relevant to conscious visual perception than assumed by that hypothesis.

According to the perception-action model, manipulations (e.g., continuous flash suppression) preventing conscious perception should nevertheless permit more processing in the dorsal than the ventral stream. However, neuroimaging studies have typically obtained no evidence that neural activity in the dorsal stream is greater than in the ventral stream when observers lack conscious awareness of visual stimuli (Hesselmann et al., 2018).

Two pathways: update

The perception-action model was originally proposed before neuroimaging and other techniques had clearly indicated the great complexity of the brain networks involved in perception and action (de Haan et al., 2018). Recent research has led to developments of the perception-action model in two main ways. First, we now know much more about the various interactions between processing in the dorsal and ventral streams. Second, there are more than two visual processing streams. Rossetti et al. (2017) show how theoretical conceptualisations of the relationship between visual perception and action have become more complex (see Figure 2.15).

We have seen that the ventral pathway is often involved in visually guided action. There is also increasing evidence the dorsal pathway is involved in visual object recognition (Freud et al., 2016). For example, patients with damage to the ventral pathway often retain some sensitivity to three-dimensional (3-D) structural object representations (Freud et al., 2017a). Zachariou et al. (2017) applied transcranial magnetic stimulation to posterior parietal cortex within the dorsal pathway to disrupt processing. TMS disrupted the holistic processing (see Glossary) of faces, suggesting the dorsal pathway is involved in face recognition.

More supporting evidence was reported by Freud et al. (2016). They studied shape processing, which is of central importance in object recognition and so should depend primarily on the ventral pathway. However, the ventral and dorsal pathways were both sensitive to shape. The observers' ability to recognise objects correlated with the shape sensitivity of regions within the dorsal pathway. Thus, dorsal path activation was of direct relevance to shape and object processing.

How many visual processing streams are there? There is evidence that actions towards objects depend on two partially separate dorsal streams (Sakreida et al., 2016; see Chapter 4). First, there is a dorso-dorsal stream (the "grasp" system) used to grasp objects rapidly. Second, there is a
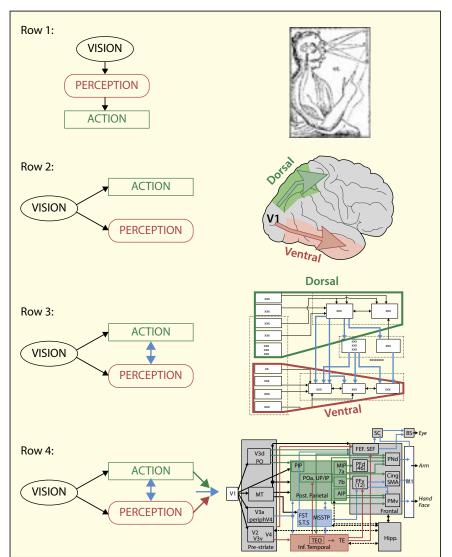
Figure 2.15
Historical developments in theories linking perception and action. Row 1: the intuitive notion that action is preceded by conscious perception. Row 2: Goodale and Milner's original two systems' theory. Row 3: interaction between the two anatomical pathways and perceptual and visual processes. Row 4: evidence that processing in primary motor cortex is preceded by interconnections between dorsal (green) and ventral (red) pathways.

From Rossetti et al. (2017). Reprinted with permission of Elsevier.

ventro-dorsal stream that makes use of memorised object knowledge and operates more slowly than the first stream.

Haak and Beckmann (2018) investigated the connectivity patterns among 22 visual areas, discovering these areas "are organised into not two but three visual pathways: one dorsal, one lateral, and one ventral" (p. 82). Their findings thus provide some support for the emphasis within the perception-action model on dorsal and ventral streams. Haak and Beckmann speculated that the new lateral pathway may "incorporate . . . aspects of vision, action and language" (p. 81).

Overall evaluation

Milner and Goodale's theoretical approach has been hugely influential. Their central assumption that there are two visual systems ("what"

and "how") is partially correct. It has received inconsistent support from research on patients with optic ataxia and visual agnosia. Earlier we discussed achromatopsia (see Glossary) and akinetopsia (see Glossary). The former condition depends on damage to the ventral pathway and the latter condition on damage to the dorsal pathway (Haque et al., 2018). As predicted theoretically, many visual illusions are much reduced in extent when observers engage in action-based performance (e.g., pointing; grasping).

What are the model's limitations? First, evidence from brain-damaged patients provides relatively weak support for it. In fact, "The idea of a double dissociation between optic ataxia and visual form agnosia, as cleanly separating visuo-motor from visual perceptual functions, is no longer tenable" (Rossetti et al., 2017, p. 130).

Second, findings based on visual illusions provide only partial support for the model. The findings generally indicate that illusory effects are greater with perceptual judgements than actions but there are many exceptions.

Third, the model exaggerates the *independence* of the two visual systems. For example, Janssen et al. (2018) reviewed research on 3-D object perception and found strong effects of the dorsal stream on the ventral stream. As de Haan et al. indicated,

> The prevailing evidence suggests that cross-talk [interactions between visual systems] is the norm rather than the exception . . . [There is] a flexible and dynamic pattern of interaction between visual processing areas in which visually processing networks may be created on-the-fly in a highly task-specific manner.
>
> (de Haan et al., 2018, p. 6)

Fourth, the notion there are only two visual processing streams is an oversimplification. Earlier on pp. 62–63 we discussed two attempts (Haak & Beckmann, 2018; Sakreida et al., 2016) to develop more complete accounts.

COLOUR VISION

Why do we have colour vision? After all, if you watch an old black-and-white movie on television you can easily understand the moving images. One reason is that colour often makes an object stand out from its surroundings making it easier to identify. Chameleons very sensibly change colour to blend in with the background, thus reducing their chances of being detected by predators.

Colour perception also helps us to recognise and categorise objects. For example, it is useful when deciding whether a piece of fruit is under- or overripe. Predictive coding (processing primarily aspects of sensory input that violate the observer's predictions) is also relevant (Huang & Rao, 2011). Colour vision allows observers to focus rapidly on any aspects of the incoming visual input (e.g., discolouring) discrepant with predictions based on ripe fruit.

There are three main qualities associated with colour:

(1) *Hue*: the colour itself and what distinguishes red from yellow or blue.
(2) *Brightness*: the perceived intensity of light.

(3) *Saturation*: this allow us to determine whether a colour is vivid or pale; it is influenced by the amount of white present.

Trichromacy theory

Retinal cones are specialised for colour vision. Cone receptors contain light-sensitive photopigment allowing them to respond to light. According to the trichromatic [three-coloured] theory, there are three kinds of receptors:

(1) One type is especially sensitive to short-wavelength light and generally responds most strongly to stimuli perceived as blue.
(2) A second type of cone receptor is most sensitive to medium-wavelength light and responds greatly to stimuli generally seen as yellow-green.
(3) A third type of cone responds most to long-wavelength light such as that reflected from stimuli perceived as orange-red.

How do we see other colours? According to the theory, most stimuli activate two or all three cone types. The colour we perceive is determined by their *relative* stimulation levels. Evolution has equipped us with three types of cones because that produces a very efficient system – we can discriminate millions of colours even with so few cone types.

Many forms of colour deficiency are consistent with trichromacy theory. Most individuals with colour deficiency have **dichromacy**, in which one cone class is missing. In red-green dichromacy (the most common form) there are abnormalities in the retinal pigments sensitive to medium or long wavelengths. Individuals with red-green dichromacy differ from intact observers in perceiving far fewer colours. However, their colour constancy (see Glossary) is almost at normal levels (Álvaro et al., 2017).

The density of cones (the retinal cells responsible for colour vision) is far higher in the fovea (see Glossary) than the periphery. However, there are enough cones in the periphery to permit accurate peripheral colour judgements if colour patches are reasonably large (Rosenholtz, 2016).

The crucial role of cones for colour vision explains the following common phenomenon: "The sunlit world appears in sparkling colour, but when night falls . . . we see the world in 50 shades of grey" (Kelber et al., 2017, p. 1). In dim light, the cones are not activated and our vision depends almost entirely on rods.

Opponent-process theory

Trichromacy theory does not explain what happens *after* activation of the cone receptors. It also fails to account for **negative afterimages**. If you stare at a square of a given colour for several seconds and then shift your gaze to a white surface, you see a negative afterimage in the complementary colour (complementary colours produce white when combined). For example, a green square produces a red afterimage, whereas a blue square produces a yellow afterimage.

Hering (1878) explained negative afterimages. He identified three types of opponent processes in the visual system. One opponent process (red-green channel) produces perception of green when responding one way and red when responding the opposite way. A second opponent process

(blue-yellow channel) produces perception of blue or yellow in the same way. The third opponent process (achromatic channel) produces the perception of white at one extreme and black at the other.

What is the value of these three opponent processes? The three dimensions associated with opponent processes provide maximally *independent* representations of colour information. As a result, opponent processes provide very efficient encoding of chromatic stimuli.

Much research supports the notion of opponent processes. First, there is strong physiological evidence for the existence of opponent cells (Shevell & Martin, 2017). Second, the theory accounts for negative afterimages (discussed above). Third, the theory claims it is impossible to see blue and yellow together or red and green, but the other colour combinations can be seen. That is precisely what Abramov and Gordon (1994) found. Fourth, opponent processes explain some types of colour deficiency. Red-green deficiency occurs when the red-green channel cannot be used, and blue-yellow deficiency occurs when individuals cannot make effective use of the blue-yellow channel.
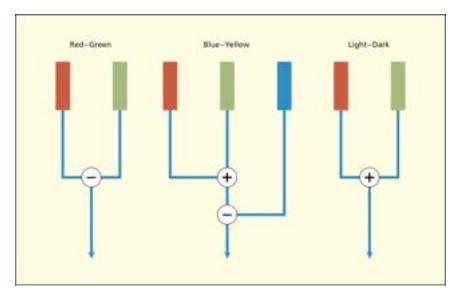
Dual-process theory

Hurvich and Jameson (1957) proposed a dual-process theory combining the ideas discussed so far. Signals from the three cones types identified by trichromacy theory are sent to the opponent cells (see Figure 2.16). There are three channels:

(1) The achromatic [non-colour] channel combines the activity of the medium- and long-wavelength cones.
(2) The blue-yellow channel represents the difference between the sum of the medium-and long-wavelength cones on the one hand and the short-wavelength cones on the other. The direction of difference determines whether blue or yellow is seen.

Figure 2.16
Schematic diagram of the early stages of neural colour processing. Three cone classes (red = long; green = medium; blue = short) supply three "channels". The achromatic (light-dark) channel receives non-spectrally opponent input from long- and medium-cone classes. The two chromatic channels receive spectrally opponent inputs to create the red-green and blue-yellow channels.

From Mather (2009). Copyright 2009 George Mather. Reproduced with permission.

(3)   The red-green channel represents the difference between activity levels in the medium- and long-wavelength cones. The direction of this difference determines whether red or green is perceived.

Overall evaluation

Dual-process theory has much experimental support. However, it is oversimplified in several ways (Shevell & Martin, 2017). First, there are complex interactions between the channels. For example, short-wavelength cones are activated even in conditions where it would be expected that only the red-green channel (involving medium- and long-wavelength cones) would be active (Conway et al., 2018). Second, the proportions of different cone types vary considerably across individuals but this typically has surprisingly little effect on colour perception. Third, the arrangement of cone types in the eye is fairly *random*. This seems odd because it presumably makes it hard for colour-opponent processes to work effectively.

More generally, much research has focused on colour perception and other research has focused on how nerve cells respond to light of different wavelengths. What has proved difficult is to relate these two sets of findings *directly* to each other. So far there is only limited convergence between psychological and physiological research (Shevell & Martin, 2017).

Colour constancy

Colour constancy is the tendency for a surface or object to be perceived as having the same colour when there are changes in the wavelengths contained in the illuminant (the light source illuminating the surface or object). Colour constancy indicates colour vision does *not* depend solely on the wavelengths of the light reflected from objects. Learn more about colour constancy on YouTube: "This is Only Red by Vsauce".

Why is colour constancy important? If we lacked colour constancy, the apparent colour of familiar objects would change dramatically when the lighting conditions altered. This would make it very hard to recognise objects rapidly and accurately.

Attaining reasonable levels of colour constancy is an impressive achievement. Look at the object in Figure 2.17. It is immediately recognisable as a blue mug even though several other colours can be perceived. The wavelengths of light depend on the mug itself, the illuminant and reflections from other objects onto the mug's surface (mutual illumination).

> **KEY TERMS**
>
> **Colour constancy**
> The tendency for an object to be perceived as having the same colour under widely varying viewing conditions.
>
> **Illuminant**
> A source of light illuminating a surface or object.
>
> **Mutual illumination**
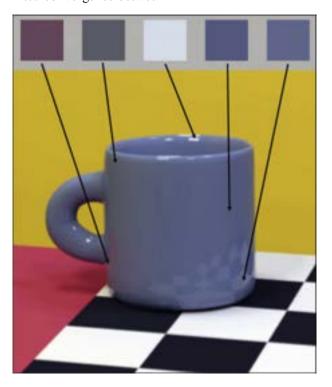> The light reflected from the surface of an object impinges on the surface of a second object.

Figure 2.17
Photograph of a mug showing enormous variation in the properties of the reflected light across the mug's surface. The patches at the top of the figure show image values from the locations indicated by the arrows.

From Brainard and Maloney (2011). Reprinted with permission of the Association for Research in Vision and Ophthalmology.

How good is colour constancy?

Colour constancy is often reasonably good. For example, Granzier et al. (2009a) assessed colour constancy for six similarly coloured papers in various indoor and outdoor locations differing substantially in lighting conditions. They found 55% of the papers were identified correctly. This represents good performance given the similarities among the papers and the large differences in lighting conditions.

Reeves et al. (2008) distinguished between our *subjective experience* and our *judgements* about the world. For example, as you walk towards a fire, it feels increasingly hot subjectively. However, how hot you *judge* the fire to be is unlikely to change. Reeves et al. found colour constancy with non-naturalistic (artificial stimuli) was much greater when observers judged the *objective* similarity of two stimuli seen under different illuminants than when rating their *subjective* similarity. Radonjić and Brainard's (2016) obtained similar findings with naturalistic stimuli. However, colour constancy was higher overall with naturalistic stimuli because such stimuli provided more cues to guide performance.

Estimating scene illumination

The wavelengths of light reflected from an object are greatly influenced by the illuminant (light source). High levels of colour constancy could be achieved if observers made accurate illuminant estimates. However, they often do not, especially when the illuminant's characteristics are unclear (Foster, 2011). For example, there are substantial individual differences in the perceived illuminant (and perceived colour) of the famous dress discussed in the Box.

Colour constancy should be high when illuminant estimation is accurate (Brainard & Maloney, 2011). Bannert and Bartels (2017) tested this prediction. Observers were presented with visual scenes using three different illuminants, and cues within the scenes were designed to facilitate colour constancy. Bannert and Bartels used functional magnetic resonance imaging (fMRI) to assess the neural encoding of each scene.

What did Bannert and Bartels (2017) find? Their key finding was that, "The neural accuracy of encoding the *illuminant* of a scene [predicted] the behavioural accuracy of constant colour perception" (p. 357). Thus, colour constancy was high when the illuminant was processed accurately.

Local colour contrast

Land (1986) proposed retinex theory, according to which we perceive a surface's colour by comparing its ability to reflect, short-, medium- and long-wavelength light against that of adjacent surfaces. Thus, we make use of local colour contrast. Kraft and Brainard (1999) studied colour constancy for complex visual scenes. Under full viewing conditions, colour constancy was 83% even with large changes in illumination. When local contrast could not be used, however, colour constancy dropped to 53%.

Foster and Nascimento (1994) developed Land's ideas into an influential theory based on local contrast. We can see the nature of their big

IN THE REAL WORLD: WHAT COLOUR IS "THE DRESS"?

On 7 February 2015, Cecilia Bleasdale took a photograph of the dress she intended to wear at her daughter's imminent wedding (see below) and posted it on the internet. It caused an almost immediate sensation because observers disagreed vehemently concerning the dress's colour.

What colour do you think the dress is (see Figure 2.18)? Wallisch (2017) found 59% of observers said the dress was white and gold and 27% said it was black and blue. How can we explain these individual differences? Wallisch argued the illumination of the dress is ambiguous: the upper part of the dress implies illumination by daylight whereas the lower part implies artificial illumination. Many theories predict the perceived colour of an object depends on its assumed illumination (discussed on p. 62). If so, observers assuming the dress is illuminated by natural light should perceive it as white and gold. In contrast, those assuming artificial illumination should perceive it as black and blue.

What did Wallisch (2017) find? As predicted, observers assuming the dress was illuminated by natural light were much more likely than those assuming artificial light to perceive the dress as white/gold (see Figure 2.19).

Figure 2.18
"The Dress" made famous by its appearance on the internet.
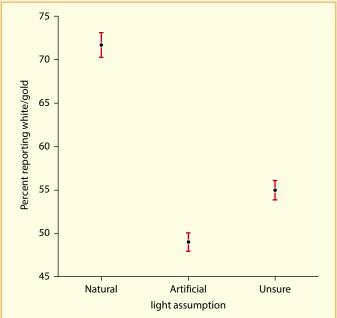From Rabin et al. (2016).

Figure 2.19
The percentage of observers perceiving "The Dress" to be white and gold depended on whether they believed it to be illuminated by natural light or by artificial light, and those who were unsure.
From Wallisch et al. (2017).

discovery through an example. Suppose there are two illuminants and two surfaces. If surface 1 led to the long-wavelength or red cones responding three times as much with illuminant 1 as illuminant 2, then the same *threefold* difference was also found with surface 2. Thus, the *ratio* of cone responses was essentially invariant across different illuminations. Thus,

cone-excitation ratios can be used to eliminate the illuminant's effects and so increase colour constancy.

Much evidence indicates cone-excitation ratios are important (Foster, 2011, 2018). For example, Nascimento et al. (2004) obtained evidence suggesting the level of colour constancy in different conditions could be predicted on the basis of cone-excitation ratios.

Foster and Nascimento's (1994) theory provides an elegant account of illuminant-independent colour constancy in simple visual environments. However, it has limited value in complex visual environments. For example, colour constancy for a given object can become harder because of reflections from other objects (see Figure 2.17) or because multiple sources of illumination are present together.

The theory is generally less applicable to natural scenes than artificial laboratory scenes. For example, the illuminant often changes more rapidly in natural scenes (e.g., clouds change shape, which influences the shadows they cast) (Nascimento et al., 2016). In addition, there are dramatic changes in the level and colour of natural illuminants over the course of the day.

In sum, cone-excitation ratios are most likely to be almost invariant, "provided that sampling is from points close together in space or time . . ., or from points separated arbitrarily but undergoing even changes in illumination" (Nascimento et al., 2016, p. 44).

Effects of familiarity

Colour constancy is influenced by our knowledge of the familiar colours of objects (e.g., bananas are yellow). Hansen et al. (2006) asked observers to view photographs of fruits and to adjust their colour until they appeared grey. There was over-adjustment. For example, a banana still looked yellowish to observers when it was actually grey, leading them to adjust its colour to a slightly bluish hue. Such findings may reflect an influence of familiar size on subjective colour perception. Alternatively, familiar colour may primarily influence observers' *responses* rather than their *perception* (e.g., our knowledge that bananas are yellow may bias us to report them as more yellow than they actually appear).

Vandenbroucke et al. (2016) investigated the above issue. Observers viewed an ambiguous colour intermediate between red and green presented on typically red (e.g., *tomato*) or green (e.g., *pine tree*) objects. Familiar colour influenced colour perception. Of most importance, neural responses in various visual areas (e.g., V4, which is much involved in colour processing) were influenced by familiar colour. Neural responses corresponded more closely to those associated with red objects when the object was typically red than when it was typically green and more closely to those found with green objects when it was typically green. Thus, familiar colour had a *direct* influence on perception early in visual processing.

Chromatic adaptation

One reason we have reasonable colour constancy is because of **chromatic adaptation** – an observer's visual sensitivity to a given illuminant decreases over time. If you stand outside after nightfall, you may be surprised by the

apparent yellowness of the artificial light in people's houses. However, this is not the case if you spend some time in a room illuminated by artificial light. Lee et al. (2012b) found some aspects of chromatic adaptation within six seconds. Such rapid adaptation increases colour constancy.

Evaluation

In view of the complexity of colour constancy, it is unsurprising the visual system adopts an "all hands on deck" approach in which several factors contribute to colour constancy. Of major importance are zone-excitation ratios that remain almost invariant across changes in illumination. In addition, top-down factors (e.g., our memory for the familiar colours of common objects) also play a role.

What are the limitations of theory and research on colour constancy? First, we lack a comprehensive theory of how the various factors *combine*. Second, most research has focused on relatively simple artificial visual environments. In contrast, "The natural world is optically unconstrained. Surface properties may vary from one point to another, and reflected light may vary from one instant to the next" (Foster, 2018, p. B192). As a result, the processes involved in trying to achieve colour constancy in more complex environments are poorly understood.

Third, more research is needed to understand *why* colour constancy depends greatly on the precise instructions given to observers. Fourth, as Webster (2016, p. 195) pointed out, "There are pronounced [individual] differences in almost all measures of colour appearance . . . the basis for these differences remains uncertain."

DEPTH PERCEPTION

A major accomplishment of visual perception is the transformation of the two-dimensional retinal image into perception of a three-dimensional world seen in depth. The construction of 3-D representations is very important if we are to pick up objects, decide whether it is safe to cross the road and so on.

Depth perception depends on numerous visual and other cues (discussed below). All cues provide *ambiguous* information and so we would be ill-advised to place total reliance on any *single* cue. Moreover, different cues often provide conflicting information. When you watch a movie, some cues (e.g., stereo ones) indicate everything you see is at the same distance. In contrast, other cues (e.g., perspective; shading) indicate some objects are closer.

In real life, depth cues are often provided by movement of the observer or objects in the visual environment and some cues are non-visual (e.g., object sounds). Here, however, the main focus will be on visual depth cues available when the observer and environmental objects are static.

Cues to depth perception are monocular, binocular and oculomotor. **Monocular cues** require only one eye but can also be used with two eyes. The fact that the world still retains a sense of depth with one eye closed indicates clearly that monocular cues exist. **Binocular cues** involve both eyes used together. Finally, **oculomotor cues** depend on sensations of

KEY TERMS 🔑

Monocular cues
Cues to depth that can be used by one eye but can also be used by both eyes together.

Binocular cues
Cues to depth that require both eyes to be used together.

Oculomotor cues
Cues to depth produced by muscular contractions of the muscles around the eye; use of such cues involves kinaesthesia (also known as the muscle sense).

KEY TERMS 🔑

Texture gradient
The rate of change of texture density from the front to the back of a slanting object.

muscular contractions of the muscles around the eye. Use of these cues involves kinaesthesia (the muscle sense).

Monocular cues

Monocular cues to depth are called *pictorial cues* because they are used by artists. Of particular importance is *linear perspective*, which artists use to create the impression of three-dimensional scenes on two-dimensional canvases. Linear perspective (based on laws of optics and geometry) is based on various principles. For example, parallel lines pointing away from us converge (e.g., motorway edges) and objects reduce in size as they recede into the distance.

Tyler (2015) argued that linear perspective is only really effective in creating a powerful 3-D effect when viewed from the point from which the artist constructed the perspective. This is typically very close to the picture as can be seen in a drawing by the Dutch artist Jan Vredeman de Vries (see Figure 2.20).

Texture is another monocular cue. Most objects (e.g., carpets; cobblestone roads) possess texture, and textured objects slanting away from us have a *texture gradient* (Gibson, 1979; see Figure 2.21). This is a gradient (rate of change) of texture density as you look from the front to the back of a slanting object with the gradient changing more rapidly for objects slanted steeply away from the observer. Sinai et al. (1998) found observers judged the distances of nearby objects better when the ground was uniformly textured than when there was a gap (e.g., a ditch) in the texture pattern.

Texture gradient is a limited cue because the perceived slant depends on the *direction* of the gradient. For reasons that are unclear, ground patterns are perceived as less slanted than equivalent ceiling or sidewall patterns (Higashiyama & Yamazaki, 2016).
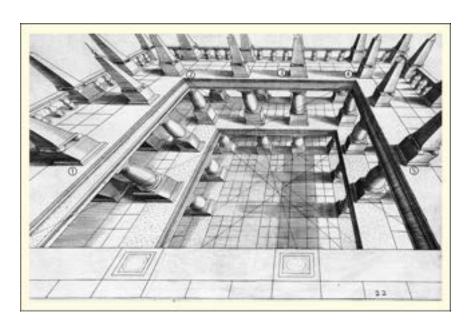

Figure 2.20
An engraving by de Vries (1604/1970) in which linear perspective creates an effective three-dimensional effect when viewed from very close but not from further away.

From Todorović (2009). Copyright 1968 by Dover Publications. Reprinted with permission from Springer.

Another monocular cue is *interposition* where a nearer object hides part of a more distant one. The strength of this cue can be seen in Kanizsa's (1976) illusory square (see Figure 2.22). There is a strong impression of a yellow square in front of four purple circles even though many of its contours are missing. This depends on processes that relatively "automatically" complete boundaries using the available information (e.g., incomplete circles).

Another useful cue is *familiar size* (discussed more fully later). If we know an object's size, we can use its retinal image size to estimate its distance. However, we can be misled. Ittelson (1951) had observers view playing cards through a peephole restricting them to monocular vision. The perceived distance was determined almost entirely by familiar size. For example, playing cards double the usual size were perceived as being twice as far away from the observers than was actually the case.

We turn now to *blur*. There is no blur at fixation point and it increases more rapidly at closer distances than ones further away. Held et al. (2012) found blur was an effective depth cue (especially at longer distances). However, observers may simply have learned to respond that the blurrier stimulus was further away. Langer and Siciliano (2015) provided minimal training and obtained little evidence blur was used as a depth cue. They argued blur provides ambiguous information: an object can appear blurred because it is in peripheral vision rather than because it is far away.

Finally, there is **motion parallax**, which involves "transformations of the retinal image that are created . . . both when the observer moves (observer-produced parallax) and when objects move with respect to the observer (object-produced parallax)" (Rogers, 2016, p. 1267). For example, when you look out of the window of a moving train, *nearby* objects appear to move in the opposite direction but *distant* objects in the same direction. Rogers and Graham (1979) found motion parallax on its own can produce accurate depth judgements. Most research demonstrating the value of motion parallax as a depth cue has used very simple random-dot displays. However, Buckthought et al. (2017) found comparable effects in more complex and naturalistic conditions.

Cues such as linear perspective, texture gradient and interposition allow observers to perceive depth even in two-dimensional displays. However, research with computer-generated two-dimensional displays has found depth is often *underestimated* (Domini et al., 2011). Such displays provide cues to flatness (e.g., binocular disparity, accommodation and vergence, all discussed on pp. 74–75) that may reduce the impact of cues suggesting depth.
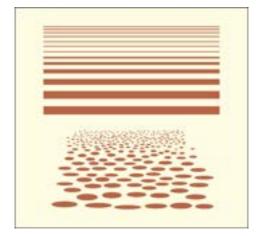

Figure 2.21
Examples of texture gradients that can be perceived as surfaces receding into the distance.
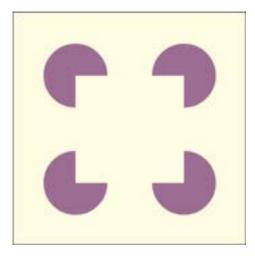From Bruce et al. (2003).

Figure 2.22
Kanizsa's (1976) illusory square.

KEY TERMS 🔑

Binocular disparity
A depth cue based on the slight disparity in the two retinal images when an observer views a scene; it is the basis for **stereopsis**.

Stereopsis
Depth perception based on the small discrepancy in the two retinal images when a visual scene is observed (**binocular disparity**).

Autostereogram
A complex two-dimensional image perceived as three-dimensional when not focused on for a period of time.

Amblyopia
A condition in which one eye sends an inadequate input to the visual cortex; colloquially known as lazy eye.

Binocular cues

Depth perception does not depend solely on monocular and oculomotor cues. It can also be achieved by binocular disparity, which is the slight difference or disparity in the images projected on the retinas of the two eyes when you view a scene (Welchman, 2016). Binocular disparity produces stereopsis (the ability to perceive the world three-dimensionally).

The great subjective advantage of binocular vision was described by Susan Barry (2009, pp. 94–132), a neuroscientist who recovered binocular vision in late adulthood:

> [I saw] palpable volume[s] of empty space . . . I could see, not just infer, the volume of space between tree limbs . . . the grape was rounder and more solid than any grape I had ever seen . . . Objects seemed more solid, vibrant, and real.

Stereopsis is very powerful at short distances. However, the disparity or discrepancy in the retinal images of objects decreases by a factor of 100 as their distance from an observer increases from 2 to 20 metres. Thus, stereopsis rapidly becomes less available at greater distances. While stereopsis provides valuable information at short distances, we must not exaggerate its importance. Bülthoff et al. (1998) found observers' recognition of familiar objects was not adversely affected when stereoscopic information was scrambled. Indeed, observers were unaware the depth information was scrambled!

Stereopsis involves *matching* features in the inputs to the two eyes. This process is fallible. For example, consider an autostereogram (a two-dimensional image containing depth information so it appears three-dimensional when viewed appropriately; the Wikipedia entry for autostereogram provides examples).

With autostereograms, the same repeating 2-D pattern is presented to each eye. If there is a dissociation of vergence and accommodation, two adjacent patterns will form an object apparently at a different depth from the background. Some individuals are better than others at perceiving 3-D objects in autostereograms because of individual differences in binocular disparity, vergence and accommodation (Gómez et al., 2012).

The most common reason for impaired stereoscopic depth perception is amblyopia (one eye exhibits poor visual acuity; also known as lazy eye). However, deficient stereoscopic depth perception can also result from damage to various cortical areas (Bridge, 2016). As Bridge concluded, intact stereoscopic depth perception requires the following: "(i) both eyes aligned and functional; (ii) control over the eye muscles and vergence to the images into alignment; (iii) initial matching of retinal images; and (iv) integration of disparity information" (p. 2).

Oculomotor cues

The pictorial cues discussed so far can all be used equally well by one-eyed individuals as by those with intact vision. Depth perception also depends on oculomotor cues based on perceiving muscle contractions

around the eyes. One such cue is **vergence** (the eyes turn inwards to focus on very close objects than those further away). Another oculomotor cue is **accommodation**. It refers to the variation in optical power produced by the thickening of the eye's lens when someone focuses on a close object.

Vergence and accommodation are both very limited. First, they only provide information about the distance of a *single* object at any given time. Second, they are both of value only when judging the distance of *close* objects. Even then, the information they provide is not very accurate.

Cue combination or integration

So far we have considered depth cues one by one. In the real world, however, we typically have access to many depth cues. How do we use these cues? One possibility is *additivity* (combining or integrating information from all cues) and another possibility is *selection* (using information from only a single cue) (Bruno & Cutting, 1988).

How could we maximise the accuracy of our depth perception? Jacobs (2002) argued we should assign more weight to reliable cues. Since cues reliable in one context may be less so in a different context, we should be *flexible* when assessing cue reliability. These considerations led Jacobs to propose two hypotheses:

(1) Less ambiguous cues (i.e., those providing consistent information) are regarded as more reliable than more ambiguous ones.
(2) A cue is regarded as reliable if inferences based on it are consistent with those based on other available cues.

Other theoretical approaches resemble that of Jacobs (2002). For example, Rohde et al. (2016, p. 36) discuss Maximum Likelihood Estimation, which is "a rule used . . . to optimally combine redundant estimates of a variable [e.g., object distance] by taking into consideration the reliability of each estimate and weighting them accordingly". We can extend this approach to include prior knowledge (e.g., natural light typically comes from above; many familiar objects have a typical size).

Finally, there are ideal-observer models (e.g., Landy et al., 2011; Jones, 2016). Many of these models are based on the Bayesian approach (see Chapter 13), in which initial probabilities are altered by new data or information (e.g., presentation of cues). Ideal-observer models involve making assumptions about the optimal way of combining the cue and other information available and comparing that against observers' actual performance.

As we will see, experimentation has benefitted from advances in virtual reality technologies. These advances permit researchers to control visual cues very precisely, thus permitting clear-cut tests of many hypotheses.

Findings

Evidence supporting Jacobs' (2002) first hypothesis was reported by Triesch et al. (2002). Observers in a virtual reality situation tracked an object defined by colour, shape and size. On each trial, *two* attributes were unreliable or inconsistent (their values changed frequently). Observers attached

increasing weight to the reliable or consistent cue and less to the unreliable cues during each trial.

Evidence supporting Jacobs' (2002) second hypothesis was reported by Atkins et al. (2001). Observers in a virtual reality environment viewed and grasped elliptical cylinders. There were three cues to cylinder depth: texture, motion and **haptic** (relating to the sense of touch).

When the haptic and texture cues indicated the same cylinder depth but the motion cue indicated a different depth, observers made increasing use of the texture cue and decreasing use of the motion cue. When the haptic and motion cues indicated the same cylinder depth but the texture cue did not, observers increasingly relied on the motion cue rather than the texture cue. Thus, whichever visual cue correlated with the haptic cue was preferred, and this preference increased with practice.

Much research suggests observers integrate cue information according to the additivity notion: they take account of most (or all) cues but attach additional weight to more reliable ones (Landy et al., 2011). However, these conclusions are based primarily on studies involving only small conflicts in the information provided by each cue.

What happens when two or more cues are in *strong* conflict? Observers typically rely heavily (or even exclusively) on only one cue, i.e., they use the selection strategy as defined by Bruno and Cutting (1988; see p. 75). This makes sense. Suppose one cue suggests an object is 10 metres away but another cue suggests it is 90 metres away. It is probably not sensible to split the difference and decide it is 50 metres away! We use the selection strategy at the movies – perspective and texture cues produce a 3-D effect, whereas we largely ignore cues (e.g., binocular disparity) indicating everything on the screen is the same distance from us.

Relevant evidence was reported by Girshick and Banks (2009) in a study on slant perception. When there was a small conflict between the information provided by binocular disparity and texture gradient cues, observers used information from both. However, when there was a large conflict between these cues, perceived slant was determined exclusively by one cue (binocular disparity or texture gradient). Interestingly, the observers were not consciously aware of the large conflict between the cues.

Do observers combine information from different cues to produce optimal performance (i.e., accurate depth perception)? Lovell et al. (2012) compared the effects of binocular disparity and shading on depth perception. Overall, binocular disparity was the more informative cue to depth, but Lovell et al. tested the effects of making it less reliable. Information from the cues was combined optimally, with observers consistently attaching more weight to reliable cues.

Many other studies have also reported that observers' depth perception is close to optimal. However, there are several studies where observers performed less impressively (Rahnev & Denison, 2018). For example, Chen and Tyler (2015) carried out a similar study to that of Lovell et al. (2012). Observers' depth judgements were strongly influenced by shading but made very little use of binocular disparity information.

Evaluation

Much has been learned about the numerous cues observers use to estimate depth or distance. Information from different depth cues is typically combined or integrated in studies assessing depth perception. There is also evidence that one cue often dominates the others when different cues conflict strongly.

Overall, as Brenner and Smeets (2018, p. 385) concluded, "By combining the many sources of information in a clever manner people obtain quite reliable judgments that are not too sensitive to violations of the assumptions of the individual sources of depth information." More specifically, observers generally attach most weight to cues providing reliable information consistent with that provided by other cues. If a cue becomes more or less reliable over time, observers generally increase or decrease its weighting appropriately. Overall, depth perception often appears close to optimal.

What are the limitations of theory and research on cue integration? First, we typically estimate distance in real-life settings where numerous cues are present and there are no large conflicts among them. In contrast, laboratory settings often provide only a few cues and these cues sometimes provide very discrepant information. The unfamiliarity of laboratory settings may sometimes cause suboptimal performance by observers and reduce generalisation to everyday life (Landy et al., 2011).

Second, the assumption that observers process several essentially independent cues *before* integrating all the information is dubious. It may apply when observers view a very limited and artificial visual display. However, natural environments typically provide observers with very rich information. In such environments, visual processing probably depends more on a global assessment of the overall structure of the environment and less on processing of specific depth cues than usually assumed (Sedgwick & Gillam, 2017). There are also issues concerning the meaning of the word "cue". For example, "Stereopsis is not a cue. It encompasses all the ways images of a scene differ in the two eyes" (Sedgwick & Gillam, 2017, p. 81).

Third, ideal-observer models differ in the assumptions used to compute "ideal" performance and the meaning of "optimal" combining of cues in depth perception (Rahnev & Denison, 2018). Most models focus on the *accuracy* of depth-perception judgements. However, there are circumstances (e.g., presence of a fierce wild animal) where rapid if somewhat inaccurate judgements are preferable. More generally, humans focus on "computational efficiency" – our goal is to maximise reward while minimising the computational costs of visual processing (Summerfield & Li, 2018). Thus, optimality of depth-perception judgements does not depend solely on performance accuracy.

Size constancy

Size constancy is the tendency for any given object to appear the same size whether its size in the retinal image is large or small. For example, if

someone walks towards you, their retinal image increases progressively but their apparent size remains the same.

Why do we show size constancy? Many factors are involved. An object's apparent distance is especially important when judging its size. For example, an object may be judged to be large even though its retinal image is very small provided it is far away. According to the size-distance invariance hypothesis (Kilpatrick & Ittelson, 1953), perceived size is proportional to perceived distance.

Findings

Haber and Levin (2001) argued that an object's perceived size depends on *memory* of its familiar size as well as perceptual information concerning its distance. Initially, observers estimated the sizes of common objects with great accuracy from memory. Then they saw various objects at close (0–50 metres) or distant (50–100 metres) viewing range and made size judgements. Some familiar objects were almost invariant in size (e.g., *bicycle*) or of varying size (e.g., *television set*); there were also unfamiliar stimuli (e.g., *ovals*).

What findings would we expect? If familiar size is important, size judgements should be more accurate for objects of invariant size than those of variable size, with size judgements least accurate for unfamiliar objects. If distance perception is all-important (and known to be more accurate for nearby objects), size judgements should be better for all object categories at close viewing range.

Haber and Levin (2001) found that size judgements were much better with objects having an invariant size than those having a variable size (see Figure 2.23). In addition, the viewing distance had a minimal effect on size judgements. Both of these findings are contrary to predictions from the size-distance invariance hypothesis.

If size judgements depend on perceived distance, size constancy should not be found when an object's *perceived* distance differs considerably from
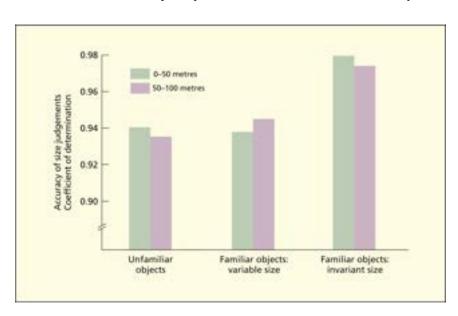
Figure 2.23
Accuracy of size judgements as a function of object type (unfamiliar; familiar variable size; familiar invariant size) and viewing distance (0–50 metres vs 50–100 metres).
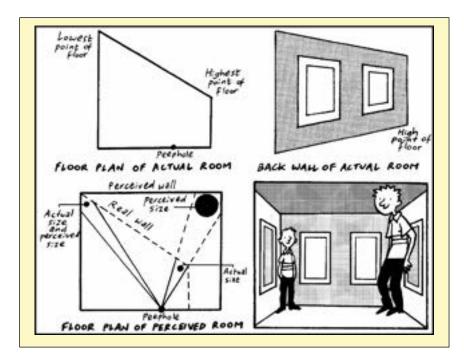Based on data in Haber and Levin (2001).

its *actual* distance. The **Ames room** (Ames, 1952; see Figure 2.24) provides a good example. It has a peculiar shape: the floor slopes and the rear wall is not at right angles to the adjoining walls. Nevertheless, the Ames room creates the same retinal image as a normal rectangular room when viewing monocularly through a peephole. The fact that one end of the rear wall is much further away from the viewer is disguised by making it much higher.

The cues suggesting the rear wall is at right angles to observers are so strong they mistakenly assume two adults standing in the corners by the

KEY TERM

Ames room
A very distorted room that nevertheless looks normal under certain viewing conditions.

KEY TERMS 🔑

Honi phenomenon
The typical apparent size changes when an individual walks along the rear wall of the **Ames room** are reduced when female observers view a man to whom they are very close emotionally.

Open-object illusion
The misperception that objects with missing boundaries are larger than objects the same size without missing boundaries.

Body size effect
An illusion in which misperception of one's own bodily size causes the perceived size of objects to be misjudged.

rear wall are at the same distance (see photograph). They thus estimate the size of the nearer adult as much greater than that of the adult further away. See the Ames room on YouTube: "Ramachandran – Ames room illusion explained".

The illusion effect with the Ames room is so great someone walking backwards and forwards in front of the rear wall seems to grow and shrink as they move! Thus, perceived distance apparently determines perceived size. However, this effect is reduced when the person walking along the rear wall is a man and the observer is a female having a close emotional relationship with him. This is known as the Honi phenomenon because it was first experienced by a woman (whose nickname was Honi) when she saw her husband in the Ames room.

Similarly dramatic findings were reported by Glennerster et al. (2006). Participants walked through a virtual-reality room as it expanded or contracted considerably. Even though they had detailed information from motion parallax and motion to indicate the room's size was changing, no participants noticed the changes! There were large errors in participants' judgements of the sizes of objects at longer distances because of their powerful expectation the size of the room would not alter.

Some evidence discussed so far has been consistent with the assumption of the size-distance invariance hypothesis that perceived size depends on perceived distance. However, many other findings are inconsistent (Kim, 2017b). For example, Kim et al. (2016) obtained size and distance estimates from observers for objects placed in various tunnels. Size and distance were perceived *independently* (i.e., depended on different factors). In contrast, the size-distance invariance hypothesis predicts that perceived size and perceived distance should depend on each other and thus should *not* be independent.

Kim (2018) obtained similar findings when observers viewed a virtual object presented stereoscopically. Their size judgements were more accurate than their distance judgements, with each judgement depending on its own information source.

More evidence inconsistent with the size-distance invariance hypothesis was reported by Makovski (2017). Participants were presented with stimuli such as those shown in Figure 2.25 on a monitor. Even though perceived distance was the same for all stimuli, "open" objects (having missing boundaries) were perceived as much larger than "closed" objects (with all boundaries intact). This is the open-object illusion in which observers extend the missing boundaries. This may resemble our common perception that open windows make a room seem larger.

Van der Hoort et al. (2011) found evidence for the body size effect, in which the size of a body mistakenly perceived to be one's own influences the perceived sizes of objects. Participants equipped with head-mounted displays connected to CCTV cameras saw the environment from the perspective of a doll (see Figure 2.26). The doll was small or large.

Van der Hoort et al. (2011) found objects were perceived as larger and further away when the doll was small than when it was large. These effects were greater when participants misperceived the body as their own (this was achieved by having the bodies of the participants and the doll

touched at the same time). Thus, size and distance perception depend partly on our lifelong experience of seeing everything from the perspective of our own body.

Tajadura-Jiménez et al. (2018) extended the above findings. Participants experienced having the body of a 4-year-old child or an adult with the body scaled down to match the height of the child's body. Object size was overestimated more in the child-body condition, indicating that object size is influenced by higher-level cognitive processes (i.e., age perception).

Evaluation

Size perception and size constancy sometimes depend on perceived distance. Some of the strongest evidence comes from research where misperceptions of distance (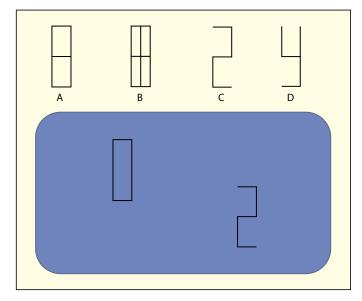e.g., in the Ames room; in virtual environments) produce systematic distortions in perceived size. However, several other factors also influence size perception. These include familiar size, one's perceived body size and whether objects do (or do not) contain missing boundaries.

What are the limitations of research and theory on size perception? First, psychologists have discovered fewer sources of information accounting for size perception than depth perception. In addition, as Kim (2017b, p. 2) pointed out, "The efficacy of the few information sources that have been identified for size perception is questionable." Second, while the size-distance invariance hypothesis remains influential, there is a "vast literature demonstrating independence of perceived size and distance" (Kim, 2018, p. 17).
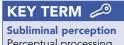
PERCEPTION WITHOUT AWARENESS: SUBLIMINAL PERCEPTION

Can we perceive aspects of the visual world without any conscious awareness we are doing so? In other words, is there such a thing as subliminal perception (stimulus perception occurring even though the stimulus is below the threshold of conscious awareness)? Common sense suggests the answer is "No". However, much research evidence suggests the answer is "Yes". However, we must use terms carefully. A thermostat responds appropriately to temperature changes and so could be said to exhibit unconscious perception!

Much important evidence has come from blindsight patients with damage to early visual cortex (V1), an area of crucial importance to

Figure 2.25

Top: stimuli presented to participants; bottom: example of the stimulus display.

From Makovski (2017).

Figure 2.26

What participants in the doll experiment could see. From the viewpoint of a small doll, objects such as a hand look much larger than when seen from the viewpoint of a large doll. This exemplifies the body size effect.

From Van der Hoort et al. (2011). Public Library of Science. With kind permission from the author.

KEY TERM

Subliminal perception
Perceptual processing occurring below the level of conscious awareness that can nevertheless influence behaviour.

Blindsight refers to patients' ability to "detect, localise, and discriminate visual stimuli in their blind field, despite denying being able to see the stimuli" Mazzi et al. (2016, p. 1).

In what follows, we initially consider blindsight patients. After that, we discuss evidence of subliminal perception in healthy individuals.

Blindsight

Many British soldiers in the First World War who had been blinded by gunshot wounds that destroyed their primary visual cortex (V1 or BA17) were treated by George Riddoch, a captain in the Royal Army Medical Corps. These soldiers responded to motion in those parts of the visual field in which they claimed to be blind. The apparently paradoxical nature of their condition was neatly captured by Weiskrantz et al. (1974), who coined the term "blindsight".

How is blindsight assessed? Various approaches have been taken but there are generally two measures. First, there is a forced-choice test in which patients guess (e.g., stimulus present or absent?) or point at stimuli they cannot see. Second, there are patients' subjective reports that they cannot see stimuli presented to their blind region. Blindsight is typically defined by an absence of self-reported visual perception accompanied by above-chance performance on the forced-choice test.

KEY TERM 🔑

Blindsight
The ability to respond appropriately to visual stimuli in the absence of conscious visual experience in patients with damage to the primary visual cortex.

IN THE REAL WORLD: BLINDSIGHT PATIENT DB 🌐

Much early research on blindsight involved a patient, DB. He was blind in the lower part of his left visual field as a result of surgery involving removal of part of his right primary visual cortex (BA17) to relieve his frequent severe migraine. DB was studied intensively by Larry Weiskrantz.

DB is one of the most thoroughly studied blindsight patients (see Weiskrantz, 2010, for a historical review). He underwent surgical removal of the right occipital cortex, including most of the primary visual cortex, to relieve very severe migraine attacks. DB could detect the presence of an object and could indicate its approximate location by pointing. He could also discriminate between moving and stationary objects and could distinguish vertical from horizontal lines. However, DB's abilities were limited – he could not distinguish between different-sized rectangles or between triangles having straight and curved sides. Such findings suggest DB processed only low-level features of visual stimuli and could not discriminate form.

We have seen DB showed some ability to perform various visual tasks. However, he reported no conscious experience in his blind field. According to Weiskrantz et al. (1974, p. 721), "When he was shown a video film of his reaching and judging orientation of lines [by presenting it to his intact visual field], he was openly astonished."

Campion et al. (1983) pointed out that DB and other blindsight patients are only partially blind. They favoured the stray-light hypothesis, according to which patients respond to light reflected from the environment onto areas of the visual field still functioning. This hypothesis implies DB should have shown reasonable visual performance when objects were presented to his blind spot (the area where the optic nerve passes through the retina). However, DB could not detect objects presented to his blind spot.

We must not exaggerate patients' preserved visual abilities. Indeed, their visual abilities in their blind field are so poor that a seeing person with comparable impairment would be legally classified as blind.

What do blindsight patients experience?

It is surprisingly hard to decide exactly what blindsight patients experience when presented with visual stimuli to their blind field. For example, the blindsight patient GY described his experiences as "similar to that of a normally sighted man who, with his eyes shut against sunlight, can perceive the direction of motion of a hand waved in front of him" (Beckers & Zeki, 1995, p. 56).

On another occasion GY was asked about his qualia (sensory experiences). He said, "That [experience of qualia] only happens on very easy trials, when the stimulus is very bright. Actually, I'm not sure I *really* have qualia then" (Persaud & Lau, 2008, p. 1048).

There is an important distinction between type-1 and type-2 blindsight. Type-1 blindsight occurs when patients have no conscious awareness of visual stimuli presented to the blind field. In contrast, type-2 blindsight occurs when patients have some residual awareness (although very different from that of healthy individuals). For example, a patient, EY, "sensed a definite pinpoint of light", although "it looks like nothing at all" (Weiskrantz, 1980). Another patient, GY, said, "You don't actually ever sense anything or see anything . . . it's more an awareness but you don't see it" (Weiskrantz, 1997). Many patients exhibit type-1 blindsight on some occasions but type-2 blindsight on others.

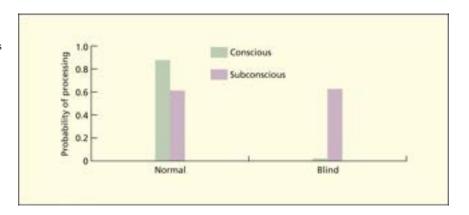## Findings: evidence for blindsight

Numerous studies have assessed the perceptual abilities of blindsight patients. Here we briefly consider three illustrative studies. As indicated already, blindsight patients often perform better when guessing an object's direction of motion than its perceptual qualities (e.g., form; colour). For example, Chabanat et al. (2019) studied a blindsight patient, SA. He was correct 98% of the time when reporting an object's direction of motion but performed at chance level when reporting its colour.

GY (discussed earlier) is a much-studied blindsight patient. He has extensive damage to the primary visual cortex in the left hemisphere. In one study (Persaud & Cowey, 2008), GY was presented with a stimulus in the upper or lower part of his visual field. On *inclusion* trials, he was instructed to report the part of the visual field to which the stimulus had been presented. On *exclusion* trials, GY was instructed to report the opposite of its actual location (e.g., "up" when it was in the lower part).

GY tended to respond with the real rather than the opposite location on exclusion and inclusion trials suggesting he had access to location information but lacked any conscious awareness of it (see Figure 2.27). In contrast, healthy individuals showed a large difference in performance on inclusion and exclusion trials indicating they had conscious access to location information.

Figure 2.27
Estimated contributions of conscious and subconscious processing to GY's performance in exclusion and inclusion conditions in his normal and blind fields.

Reprinted from Persaud and Cowey (2008). Reprinted with permission from Elsevier.

Persaud et al. (2011) manipulated the stimuli presented to GY so his visual performance was comparable in both fields. However, GY indicated conscious awareness of far more stimuli in the intact field than the blind one (43% of trials vs 3%, respectively). GY had substantially more activation in the prefrontal cortex and parietal areas to targets presented in the intact field suggesting those targets were processed much more thoroughly.

Blindsight vs degraded conscious vision

Some researchers argue blindsight patients exhibit *degraded* vision rather than a total absence of conscious awareness of "blind" field stimuli. For example, Overgaard et al. (2008) asked a blindsight patient, GR, to decide whether a triangle, circle or square had been presented to her blind field. In one experiment, GR simply responded "yes" or "no". In another experiment, Overgaard et al. used a 4-point Perceptual Awareness Scale: "clear image", "almost clear image", "weak glimpse" and "not seen".

Using the yes/no measure, GR indicated she had not seen the stimulus on 79% of trials. However, she identified it correctly 46% of the time. These findings suggest the presence of type-1 blindsight. With the 4-point scale, in contrast, GR was correct 100% of the time when she had a clear image, 72% of the time when her image was almost clear, 25% when she had a weak glimpse and 0% when the stimulus was not seen. If the "clear image" and "almost clear image" data are combined, GR claimed awareness of the stimulus on 54% of trials, on 83% of which she was correct. Thus, the use of a sensitive method (the 4-point scale) suggested much of GR's apparent blindsight reflected degraded conscious vision.

Ko and Lau (2012) argued blindsight patients have more conscious visual experience than usually assumed. Their key assumption was as follows: "Blindsight patients may use an unusually conservative criterion for detection, which results in them saying 'no' nearly all the time to the question of 'do you see something?'" (Ko & Lau, 2012, p. 1402). This excessive caution may occur in part because damage to the prefrontal cortex impairs their ability to set the criterion for visual detection appropriately. Their excessive conservatism or caution may explain why the reported visual experience of blindsight patients is so discrepant from their forced-choice perceptual performance.

Ko and Lau's (2012) theoretical position is supported by Overgaard et al.'s (2008) finding (discussed on p. 84) that blindsight patients were very reluctant to admit to having seen stimuli presented to their blind field. They also cited research supporting their assumption that blindsight patients often have prefrontal damage.

Mazzi et al. (2016) carried out a study resembling that of Overgaard et al. (2008) on another blindsight patient, SL, showing no activity in the primary visual cortex (V1). SL decided which of two features (e.g., red or green colour) was present in a stimulus. When she indicated whether she had seen the stimulus or was merely guessing, her guessing performance was significantly above chance suggestive of type-1 blindsight. However, when she indicated her awareness using the 4-point Perceptual Awareness Scale, her visual performance was at chance level when she reported no awareness of the stimulus. These findings suggest an absence of blindsight. The title of Mazzi et al.'s article provides the take-home message: "Different measures tell a different story" (p. 1).

What can we conclude? Overgaard and Mogensen (2015, p. 37) argued that "rudimentarily analysed visual information is available in blindsight" but typically does not lead to conscious awareness. However, such information can produce conscious awareness if the patient uses much effort and top-down control.

Two findings support this approach. First, blindsight patients generally do not regard their experiences as "visual" because they differ so much from normal visual perception. Second, there is much evidence (Overgaard & Mogensen, 2015) that blindsight patients show enhanced visual performance (and sometimes subjective awareness) after training. This occurs because they make increasingly effective use of the rudimentary visual information available to them.

Blindsight and the brain

As indicated above, the main brain damage in blindsight patients is to V1 (the primary visual cortex). As we saw earlier on p. 47 in the chapter, visual processing typically proceeds from V1 (BA17) to other brain areas (e.g., V2, V3, V4; see Figure 2.4). Of importance, stimuli presented to the "blind" field often produce some activation in these other brain areas. However, this activation is not associated with visual awareness in blindsight patients.
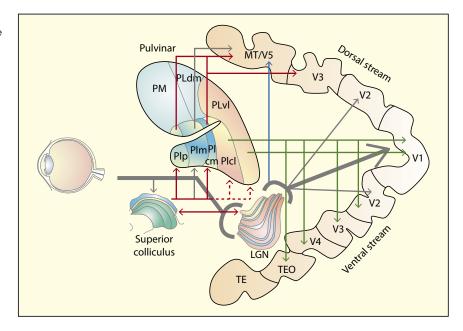
On p. 48 in the chapter we discussed research by Hurme et al. (2017) designed to clarify the role of V1 (the primary visual cortex) in the visual perception of healthy individuals. Transcranial magnetic stimulation applied to the primary visual cortex to reduce its efficiency disrupted unconscious and conscious vision. In a similar study, Hurme et al. (2019) found TMS applied to V1 prevented conscious and unconscious motion perception in healthy individuals.

In view of the above findings, how is it that many blindsight patients provide evidence of unconscious visual and motion processing? Part of the answer lies within the lateral geniculate nucleus of the thalamus, an intermediate relay station between the eye and V1 (see Figure 2.28). Ajina et al. (2015) divided patients with V1 damage into those with or without blindsight. *All* those with blindsight had intact connections between LGN and

Figure 2.28

The areas of most relevance to blindsight are the lateral geniculate nucleus (LGN) and middle temporal visual area (MT/V5). The structure close to the LGN is the pulvinar.

From Tamietto and Morrone (2016).

MT/V5 (blue arrow in the figure) whereas those connections were impaired in patients without blindsight. This finding is important given the crucial importance of MT/V5 for motion perception.

Celeghin et al. (2019) reported a meta-analysis (see Glossary) providing a fuller account of the brain areas associated with patients' visual processing. They identified 14 such areas. Some of these areas (e.g., the LGN; the pulvinar) are critical for non-conscious motion perception, whereas others (e.g., superior temporal gyrus; amygdala) are involved in non-conscious emotion processing. Overall, the meta-analysis strongly suggested that blindsight typically consists of *several* non-conscious visual abilities rather than one. Of interest, prefrontal areas (e.g., dorsolateral prefrontal cortex) often associated with conscious visual perception (see Chapter 16) were *not* activated during visual processing by blindsight patients. These findings support the view that visual processing in these patients is typically unaccompanied by conscious experience.

Finally, Celeghin et al. (2019) discussed evidence that there is substantial *reorganisation* of brain connectivity in many blindsight patients following damage to V1 (primary visual cortex). For example, consider the blindsight patient, GY, whose left V1 was destroyed. He has nerve fibre connections between the undamaged right lateral geniculate nucleus and the *contralesional* (opposite side of the body) visual motion area MT/V5 (Bridge et al., 2008) – connections not present in healthy individuals. Such reorganisation helps to explain the visual abilities displayed by blindsight patients.

Evaluation

Much has been learned about the nature of blindsight. First, two main types of blindsight have been identified. Second, evidence for the existence

of blindsight often depends on the precise measure of visual awareness used. Third, brain connections important in blindsight (e.g., between the lateral geniculate nucleus and MT/V5) have been discovered. Fourth, the visual abilities of many blindsight patients probably depend on the reorganisation of connections within the brain following damage to the primary visual cortex. Fifth, the assumption that visual processing is rudimentary in blindsight patients explains many findings. Sixth, research on blindsight has shed light on the many visual pathways that bypass V1 but whose functioning can be overshadowed by pathways involving V1 (Celeghin et al., 2019).

What are the limitations of research in this area? First, there are considerable differences among blindsight patients with several apparently possessing some conscious visual awareness in their allegedly blind field. Second, many blindsight patients have more conscious visual experience in their "blind" field than appears from yes/no judgements about stimulus awareness. This probably happens because they are excessively cautious about claiming to have seen a stimulus (Mazzi et al., 2016; Overgaard et al., 2008).

Third, the extent to which blindsight patients have degraded vision remains controversial. Fourth, the existence of reorganisation within the brain in blindsight patients (e.g., Bridge et al., 2008) may limit the applicability of findings from such patients to healthy individuals.

Subliminal perception

In research on subliminal perception in visually intact individuals, a performance measure of perception (e.g., enhanced speed or accuracy of responding) is typically compared with an awareness measure. We can distinguish between subjective and objective measures of awareness: *subjective* measures involve self-reports concerning observers' awareness, whereas *objective* measures involve forced-choice responses (e.g., did the stimulus belong to category A or B?) (Hesselmann, 2013). As Shanks (2017, p. 752) argued, "Unconscious processing [subliminal perception] is inferred when above-chance performance is combined with null awareness."

For example, Naccache et al. (2002) had observers decide rapidly whether a visible target digit was smaller or larger than 5. Unknown to them, an invisible masked digit on the same side of 5 as the target (congruent) or the other side (incongruent) was presented immediately before the target. There were two main findings. First, responses to the target digits were faster on congruent than incongruent trials (performance measure). Second, no participants reported seeing any masked digits (subjective awareness measure) and their performance was at chance level when guessing whether masked digits were below or above 5 (objective awareness measure). These findings suggested the existence of subliminal perception.

Findings

Persaud and McLeod (2008) tested the notion that only information perceived with awareness can control our actions. They presented the letter "b" or "h" for 10 ms (short interval) or 15 ms (long interval). In the key condition, participants were instructed to respond with the letter *not* presented.

For example, if they were aware "b" had been presented, they would say "h". The rationale was that only participants consciously aware of the letter could inhibit saying it.

Persaud and McLeod (2008) found participants responded correctly with the non-presented letter on 83% of long-interval trials indicating reasonable conscious awareness. In contrast, participants responded correctly on only 43% of short-interval trials (significantly below chance) suggesting some stimulus processing but an absence of conscious awareness.

An important issue is whether perceptual awareness is all-or-none (i.e., present or absent) or graded (i.e., varying in extent). Evidence suggesting it is graded was reported by Sandberg et al. (2010). One of four shapes was presented very briefly followed by masking. Observers made a behavioural response (deciding which shape had been presented) followed by one of three subjective measures: (1) clarity of perceptual experience (the Perceptual Awareness Scale); (2) confidence in their decision; and (3) wagering variable amounts of money on having made the correct decision.

What did Sandberg et al. (2010) find? First, above-chance task performance sometimes occurred without reported awareness with all three subjective measures. Second, the Perceptual Awareness Scale predicted performance better than the other measures, probably because it was the most sensitive measure of conscious experience.

The partial awareness hypothesis (Kouider et al., 2010) potentially explains graded perceptual experience. According to this hypothesis, perceptual awareness can be limited to low-level features (e.g., colour) while excluding high-level features (e.g., face identity). Supportive evidence was reported by Gelbard-Sagiv et al. (2016) with faces coloured blue or green. They used continuous flash suppression (CFS): a stimulus presented to one eye cannot be seen consciously when rapidly changing patterns are presented to the other eye. Observers often had conscious awareness of the colour of faces they could not identify.

Koivisto and Grassini (2016) presented stimuli to one of four locations. Observers then made a forced-choice responses concerning the stimulus location and rated their subjective visual awareness of the stimulus on a 3-point version of the Perceptual Awareness Scale (discussed above). Of central importance was the no-awareness category (i.e., "*I did not see any stimulus*"). The finding that observers were correct on 38% of trials associated with no awareness (chance performance = 25%) was apparent evidence for subliminal perception.

However, there is an alternative explanation. According to Koivisto and Grassini (2016, p. 241), the above finding occurred mainly when "observers were very weakly aware of the stimulus, but behaved conservatively and claimed not having seen it". This conservatism is known as *response bias*. Two findings supported this explanation. First, nearly all the observers showed response bias on no-awareness trials (see Figure 2.29).

Second, Koivisto and Grassini (2016) used event-related potentials. The N200 (a negative wave 200 ms after stimulus presentation) is typically substantially larger for stimuli associated with awareness. Of key importance, the N200 was greater on no-awareness correct trials than
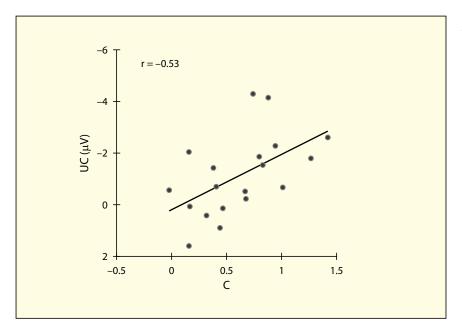
no-awareness incorrect trials for observers with high-response bias but not those with low-response bias (see Figure 2.29).

In sum, Koivisto and Grassini (2016) provided a coherent explanation for the finding that visual performance was well above chance on no-awareness trials. Observers often had weak conscious awareness on correct no-awareness trials (indicated by the N200 findings). Such weak conscious awareness occurred most frequently among those most biased against claiming to have seen the stimulus.

Neuroimaging research has consistently shown that stimuli of which the observers are unaware nevertheless produce activation in several brain areas. In one study (Rees, 2007), activation was assessed in brain areas associated with face processing and with object processing while invisible pictures of faces or houses were presented. The identity of the picture (face vs house) could be predicted with almost 90% accuracy from patterns of brain activation. Thus, subliminal stimuli can be processed reasonably thoroughly by the visual system.

Research focusing on differences in brain activation between conditions where there is (or is not) conscious perceptual awareness is discussed thoroughly in Chapter 16. Here we will mention two major findings. First, there is much less integrated or synchronised brain activation when there is no conscious perceptual awareness (e.g., Godwin et al., 2015; Melloni et al., 2007). Second, activation of areas within the prefrontal cortex (involved in integrating brain activity) is much greater for consciously perceived visual stimuli than those not consciously perceived (e.g., Gaillard et al., 2009; Godwin et al., 2015).

What do these findings mean? They strongly suggest processing is predominantly limited to low-level features (e.g., colour; motion) when stimuli are not consciously perceived, which is consistent with the partial awareness hypothesis (Kouider et al., 2010).

Evaluation

Evidence for unconscious or subliminal perception has been reported in numerous studies using numerous tasks. Some evidence is behavioural (e.g., Naccache et al., 2002; Persaud & McLeod, 2008) and some is based on patterns of brain activity (e.g., Melloni et al., 2007; Rees, 2007). The latter line of research suggests there can be considerable low-level processing of visual stimuli in the absence of conscious visual awareness. In spite of limitations of research in this area (see below), there is reasonably strong evidence for subliminal perception.

What are the limitations of research on subliminal perception? First, measures of conscious awareness vary in sensitivity. As a consequence, it is relatively easy for researches to apparently demonstrate the existence of subliminal perception by using an insensitive measure (Rothkirch & Hesselmann, 2017).

Second, many researchers focus on observers whose verbal reports show a lack of awareness. That would be appropriate if such reports were totally reliable. However, such reports are somewhat unreliable meaning that some of them would report awareness if they provided a second verbal report (Shanks, 2017). In addition, limitations of attention and memory may sometimes cause observers' reports to omit some of their conscious experience from verbal reports (Lamme, 2010).

Third, many claimed demonstrations of subliminal perception are flawed because of the typical failure to consider and/or control response bias (Peters et al., 2016). In essence, observers with response bias may claim to have no conscious awareness of visual stimuli when they actually have partial awareness (Koivisto and Grassini, 2016).

Fourth, Breitmeyer (2015) identified 24 different methods used to make visual stimuli inaccessible to visual awareness. Neuroimaging and other techniques have been used to estimate the amount of unconscious processing associated with each method. Some methods (e.g., object-substitution masking: a visual stimulus is replaced by dots surrounding it) are associated with much more unconscious processing than others (e.g., binocular rivalry, see Glossary). Of key relevance here, the likelihood of obtaining evidence for subliminal perception depends substantially on the method used to suppress visual awareness.

CHAPTER SUMMARY 📝

- **Vision and the brain.** In the retina, there are cones (specialised for colour vision) and rods (specialised for motion detection). The retina-geniculate-striate pathway between the eye and cortex is divided into partially separate P and M pathways. The dorsal stream (associated with the M pathway) terminates in the parietal cortex and the ventral stream (associated with the P pathway) terminates in the inferotemporal cortex. There are numerous interactions between the two pathways and the two streams.

According to Zeki's functional specialisation theory, different cortical areas are specialised for different visual functions (e.g., form; colour; motion). This is supported by findings from patients with selective visual deficits (e.g., achromatopsia; akinetopsia). However, much visual processing depends on large brain networks rather than specific areas and Zeki de-emphasised the importance of top-down (recurrent) processing. It remains unclear how we integrate the outputs of different visual processes (the binding problem). However, selective attention, synchronised neural activity and combining bottom-up (feedforward) processing and top-down (recurrent) processing all play a role.

• **Two visual systems: perception-action model.** Milner and Goodale identified a vision-for-perception system based on the ventral stream and a vision-for-action system based on the dorsal stream. There is limited (and inconsistent) support for the predicted double dissociation between patients with optic ataxia (damage to the dorsal stream) and visual form agnosia (damage to the ventral stream). Illusory effects found when perceptual judgements are made (ventral stream) are often much reduced when grasping or pointing responses are used (dorsal stream).

   However, such findings are often hard to interpret, and visually guided action often relies more on the ventral stream than acknowledged theoretically. More generally, the two visual systems interact with each other much more than previously assumed and there are probably more than two visual pathways.

• **Colour vision.** Colour vision helps us detect objects and make fine discriminations among them. According to dual-process theory, there are three types of cone receptors and three types of opponent processes (green-red; blue-yellow; white-black). This theory explains negative afterimages and colour deficiencies but is oversimplified. Colour constancy occurs when a surface's perceived colour remains the same when the illuminant changes. Colour constancy is influenced by our ability to assess the illuminant accurately; local colour contrast; familiarity of object colour; chromatic adaptation; and cone-excitation ratios. Most theories are more applicable to colour vision with simple artificial stimuli than complex objects in the natural world.

• **Depth perception.** There are numerous monocular cues to depth (e.g., linear perspective; texture; familiar size) plus oculomotor and binocular cues. Cues are sometimes combined additively in depth perception. However, more weight is generally given to reliable cues than unreliable ones with weightings changing if a cue's reliability alters. However, one cue often dominates all others when different cues conflict strongly. It is often assumed that observers generally combine cues near-optimally, but it is hard to

define "optimality". The assumption that observers process several independent cues prior to integrating all the information is probably wrong in natural environments providing rich information about overall environmental structure.

Size perception is sometimes strongly influenced by perceived distance as predicted by the size-distance invariance hypothesis. However, the impact of familiar size on depth perception cannot be explained by that hypothesis. More generally, perceived size and perceived distance often depend on different factors.

- **Perception without awareness: subliminal perception.** Patients with extensive damage to V1 sometimes suffer from blindsight. This is a condition involving some ability to respond to visual stimuli in the absence of normal conscious visual awareness (especially motion detection). There is no conscious awareness in type-1 blindsight but some residual awareness in type-2 blindsight. Blindsight patients are sometimes excessively cautious when reporting their conscious experience. The visual abilities of some blindsight patients probably depend on reorganisation of brain connections following brain damage.

  There is much behavioural and neuroimaging evidence for subliminal perception in visually intact individuals. However, there are problems of interpretation caused by insensitive (and unreliable) measures of self-reported awareness. Some observers may show apparent subliminal perception because they have a response bias leading them to claim no conscious awareness of visual stimuli of which they actually have limited awareness.

FURTHER READING

Brenner, E. & Smeets, J.B.J. (2018). Depth perception. In J.T. Serences (ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. 2: Sensation, Perception, and Attention* (4th edn; 385–414). New York: Wiley. The authors provide a comprehensive account of theory and research on depth perception.

de Haan, E.H.F., Jackson, S.R. & Schenk, T. (2018). Where are we now with "what" and "how"? *Cortex, 98*, 1–7. Edward de Haan and his colleagues provide an evaluation of the perception-action model

Goldstein, E.B. & Brockmole, J. (2017). *Sensation and Perception* (10th edn). Boston: Cengage. There is coverage of key areas within visual perception in this introductory textbook.

Naccache, L. (2016). Chapter 18: Visual consciousness: A "re-updated" neurological tour. *Neurology of Consciousness* (2nd edn; pp. 281–295). Lionel Naccache provides a theoretical framework within which to understand blindsight and other phenomena associated with visual consciousness.

Shanks, D.R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review, 24*, 752–775. David Shanks discusses some issues relating to research claiming to provide evidence for subliminal perception.

Tang, F. (2018). Foundations of vision. In J.T. Serences (ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. 2: Sensation, Perception, and Attention* (4th edn; pp. 1–62). New York: Wiley. Frank Tong provides a comprehensive account of the visual system and its workings.

Witzel, C. & Gegenfurtner, K.R. (2018). Colour perception: Objects, constancy, and categories. *Annual Review of Vision Science*, *4*, 475–499. Christoph Witzel and Karl Gegenfurtner discuss our current knowledge of colour perception.

Chapter

3

Object and face recognition

INTRODUCTION

Tens of thousands of times every day we identify or recognise objects in the world around us. At this precise moment, you are looking at this book. If you raise your eyes, perhaps you can see a wall and windows. Object recognition typically happens so effortlessly it is hard to believe it is actually a complex achievement. Evidence of its complexity comes from numerous unsuccessful attempts to program computers to "perceive" the environment. However, computer programs that are reasonably effective at recognising complicated two-dimensional patterns have been developed.

Why is visual perception so complex? First, objects often overlap and so we must decide where one object ends and the next one starts. Second, numerous objects (e.g., chairs; trees) vary enormously in their visual properties (e.g., colour; size; shape) and so it is hard to assign such diverse stimuli to the same category. Third, we recognise objects almost regardless of orientation (e.g., we can easily identify a plate that appears elliptical).

We can go beyond simply identifying objects. For example, we can generally describe what an object would look like from different angles, and we also know its uses and functions. All in all, there is much more to object recognition than might be supposed (than meets the eye?).

What is discussed in this chapter? The overarching theme is to unravel the mysteries associated with recognising *three-dimensional* objects. However, we initially discuss how *two-dimensional* patterns are recognised. Then the focus shifts to how we decide which parts of the visual world belong together and thus form separate objects. This is a crucial early stage in object recognition. After that, general theories of object recognition are evaluated against the available neuroimaging and behavioural evidence.

Face recognition (vitally important in our everyday lives) differs in important ways from object recognition. Accordingly, we discuss face recognition in a separate section. Finally, we consider whether the processes involved in visual imagery resemble those involved in visual perception.

Other issues relating to object recognition (e.g., depth perception; size constancy) were discussed in Chapter 2.

PATTERN RECOGNITION

We spend much of our time (e.g., when reading) engaged in **pattern recognition** – the identification or categorisation of two-dimensional patterns. Much research has considered how alphanumeric patterns (alphabetical and numerical symbols) are recognised. A key issue is the *flexibility* of the human perceptual system (e.g., we can recognise the letter "A" rapidly and across wide variations in orientation, typeface, size and writing style).

Patterns can be regarded as consisting of a set of specific features or attributes (Jain & Duin, 2004). For example, the key features of the letter "A" are two straight lines and a connecting cross-bar. An advantage of this feature-based approach is that visual stimuli varying greatly in size, orientation and minor details can be identified as instances of the same pattern.

Many feature theories assume pattern recognition involves processing specific features followed by more global or general processing to integrate feature information. However, Navon (1977) argued global processing often *precedes* more specific processing. He presented observers with stimuli such as the one shown in Figure 3.1. On some trials, they decided whether the large letter was an "H" or an "S"; on others, they decided whether the small letters were Hs or Ss.

Navon (1977) found performance speed with the small letters was greatly slowed when the large letter differed from the small letters. However, decision speed with the large letters was uninfluenced by the nature of the small letters. Navon concluded we often see the forest (global structure) before the trees (features).

There are limitations with Navon's (1977) research and conclusions. First, Dalrymple et al. (2009) found performance was faster at the level of the small letters than the large letter when the small letters were relatively large and spread out. Thus, attentional processes influence performance.

Second, Navon failed to distinguish adequately between *encoding* (neuronal responses triggered by visual stimuli) and *decoding* (conscious perception of those stimuli) (Ding et al., 2017). Encoding typically progresses from lower-level representations of simple features to higher-level representations of more complex features (Felleman & Van Essen, 1991). In contrast, Ding et al. (2017, p. E9115) found, "The brain prioritises decoding of higher-level features because they are . . . more invariant and categorical, and thus easier to . . . maintain in noisy working memory." Thus, Navon's (1977) conclusions may be more applicable to visual decoding (conscious perception) than the preceding internal neuronal responses.

KEY TERM

Pattern recognition
The ability to identify or categorise two-dimensional patterns (e.g., letters; fingerprints).

COMPANION WEBSITE
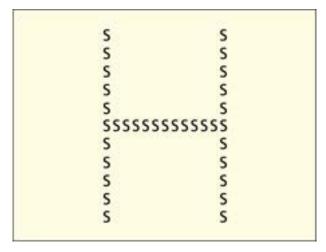
Interactive exercise:
Navon

Figure 3.1
The kind of stimulus used by Navon (1977) to demonstrate the importance of global features in perception.

Feature detectors

If presentation of a visual stimulus leads to detailed processing of its basic features, we should be able to identify cortical cells involved in such processing. Hubel and Wiesel (1962) studied cells in parts of the occipital cortex involved in visual processing. Some cells responded in two different ways to a spot of light depending on which part of the cell was affected:

(1) An "on" response with an increased rate of firing when the light was on.
(2) An "off" response with the light causing a decreased rate of firing.

Hubel and Wiesel (e.g., 1979) discovered two types of neuron in the primary visual cortex: simple cells and complex cells. Simple cells have "on" and "off" rectangular regions. These cells respond most to dark bars in a light field, light bars in a dark field, or straight edges between areas of light and dark. Any given cell responds strongly only to stimuli of a particular orientation and so its responses could be relevant to feature detection.

Complex cells resemble simple cells in responding maximally to straight-line stimuli in a particular orientation. However, complex cells have large receptive fields and respond more to moving contours. Each complex cell is driven by several simple cells having the same orientation preference and closely overlapping receptive fields (Alonso & Martinez, 1998). There are also end-stopped cells. Their responsiveness depends on stimulus length and orientation. In sum, Hubel and Wiesel envisaged, "A hierarchically organised visual system in which more complex visual features are built (bottom-up) from more simple ones" (Ward, 2015, p. 111).

Hubel and Wiesel's account is limited in several ways:

(1) The cells they identified provide *ambiguous* information because they respond comparably to different stimuli (e.g., a horizontal line moving rapidly and a nearly horizontal line moving slowly). Observers must combine information from numerous neurons to remove ambiguities.
(2) Neurons differ in their responsiveness to different spatial frequencies and several phenomena in visual perception depend on this differential responsiveness (discussed on pp. 104–105).
(3) As Schulz et al. (2015, p. 1022) pointed out, "The responses of cortical neurons [in the primary visual cortex] to repeated presentations of a stimulus are highly variable." This variability complicates pattern recognition.
(4) Pattern recognition and object recognition depend on top-down processes triggered by expectations and context (e.g., Goolkasian & Woodberry, 2010; discussed on pp. 111–116) as well as on the bottom-up processes emphasised by Hubel and Wiesel.

PERCEPTUAL ORGANISATION

Our visual environment is typically complex and confusing with many objects overlapping others, thus making it hard to achieve perceptual segregation of visual objects. How this is done was first studied systematically

IN THE REAL WORLD: HOW CAN WE DISCOURAGE SPAMMERS?

Virtually everyone has received a substantial amount of spam (unwanted emails). Spammers use bots (robots running automated tasks over the internet) to send emails to thousands of individuals for various money-making purposes (e.g., fake sweepstake entries).

A CAPTCHA (Completely Automated Turing test to tell Computers and Humans Apart) is commonly used to discourage spammers. The intention is to ensure a website user is human by providing a test humans can solve but automated computer-based systems cannot. The CAPTCHA in Figure 3.2 is typical in consisting of distorted characters connected together horizontally. In principle, the study of CAPTCHAs can shed light on the strengths of human pattern recognition.

> **KEY TERM**
>
> **CAPTCHA**
> A Completely Automated Turing Test to tell Computers and Humans Apart involving distorted characters connected together is often used to establish that the user of an internet website is human rather than an automated system.

Computer programs to solve CAPTCHAs generally involve a *segmentation* phase to locate the characters followed by a *recognition* phase where each character is identified. Many computer programs can recognise individual characters even when very distorted but their performance is much worse at segmenting connected characters. Overall, the performance of most computer programs at solving CAPTCHAs was poor until fairly recently.

Nachar et al. (2015) devised a computer program focusing on edge corners (an edge corner is the intersection of two straight edges). Such corners are relatively unaffected by the distortions and overlaps of characters found in CAPTCHAs. Nachar et al.'s approach proved successful, allowing them to solve 57% of CAPTCHAs resembling the one shown in Figure 3.2.

Figure 3.2
The CAPTCHA used by Yahoo.
From Gao et al. (2012).

There are two take-home messages. First, the difficulties encountered in devising computer programs to solve CAPTCHAs indicate humans have excellent pattern-recognition abilities. Second, edge corners provide an especially valuable source of information in pattern recognition. Of relevance, successful camouflage in many species depends heavily on markings that break up an animal's edges, making it less visible (Webster, 2015).

IN THE REAL WORLD: FINGERPRINTING

An important form of real-world pattern recognition involves experts matching a criminal's fingerprints (latent print) against stored fingerprint records. Automatic fingerprint identification systems (AFISs) scan huge databases. This typically produces a small number of possible matches to the fingerprint obtained from the crime scene ranked by similarity to the criminal's fingerprint. Experts then decide which database fingerprint (if any) matches the criminal's.

We might imagine experts are much better at fingerprint matching than novices because their analytic (slow, deliberate) processing is superior. However, Thompson and Tangen (2014) found experts greatly outperformed novices when pairs of fingerprints were presented for only 2 seconds, forcing them to rely heavily on non-analytic (fast and relatively "automatic") processing. However, when fingerprint pairs were presented for 60 seconds, experts showed a greater performance

improvement than novices (19% vs 7%, respectively). Thus, experts have superior analytic *and* non-analytic processing.

According to signal-detection theory, experts may surpass novices in their ability to *discriminate* between matching and non-matching prints. Alternatively, they may simply have a more lenient *response bias* than novices. If so, they would tend to respond "match" to every pair of prints. Good discrimination is associated with many "hits" (responding "match" on match trials) plus a low false-alarm rate (not responding "match" on non-match trials). In contrast, a lenient response criterion is associated with many false alarms.

Thompson et al. (2014) found novices made false alarms on 57% of trials on which two prints were similar but did not match, whereas experts did so on only 1.65% of trials. Thus, experts have a much more conservative response criterion as well as much better discrimination between matching and non-matching prints.

It is often assumed expert fingerprint identification is very accurate. However, experts listing the minutiae (features) on fingerprints on two occasions showed total agreement between their assessments only 16% of the time (Dror et al., 2012). Nevertheless, experts are much less likely than non-experts to decide incorrectly that two fingerprints from the same person are from different individuals (Champod, 2015).

Fingerprint identification is often complex. As an example, try to decide whether the fingerprints in Figure 3.3 come from the same person. Four fingerprinting experts said the fingerprint on the right was from the same person as the one on the left (Ouhane Daoud, the bomber involved in the terrorist attack in Madrid on 11 March 2004). In fact, the one on the right came from Brandon Mayfield, an American lawyer who was falsely arrested.

Experts' mistakes are often due to the incompleteness of the fingerprints found at crime scenes. However, top-down processes also contribute. Experts' errors often involve forensic confirmation bias: "an individual's pre-existing beliefs, expectations, motives, and situational context influence the collection, perception, and interpretation of evidence" (Kassin et al., 2013, p. 45).

Dror et al. (2006) found evidence of forensic confirmation bias. Experts were asked to judge whether two fingerprints matched having been told, incorrectly, that they were the ones mistakenly matched by the FBI as the Madrid bomber. In fact, these experts had judged these fingerprints to be a clear and definite match several years earlier. The misleading information provided led 60% of them to judge the prints to be definite non-matches! Thus, top-down processes triggered by contextual information can distort fingerprint identification.

Langenburg et al. (2009) studied the effects of context (e.g., alleged conclusions of internationally respected experts) on fingerprint identification. Experts and non-experts were both influenced by contextual information (and so showed confirmation bias). However, non-experts were influenced more.

The above studies on confirmation bias manipulated context very directly and explicitly. Searston et al. (2016) found a more subtle context effect based on familiarity. Novice participants were presented initially with a series of cases and fingerprint pairs and given feedback as to whether the fingerprints matched or not. Then they were presented with cases very similar to those seen previously

Figure 3.3
The FBI's mistaken identification of the Madrid bomber. The fingerprint from the crime scene is on the left. The fingerprint of the innocent suspect (positively identified by various fingerprint experts) is on the right.

From Dror et al. (2006). Reprinted with permission from Elsevier.

and decided whether the fingerprint pairs matched. The participants exhibited response bias during the second part of the experiment: their decisions (i.e., match or no-match) tended to correspond to the correct decisions associated with similar (but not identical) cases encountered earlier.

In sum, experts typically outperform novices at fingerprint matching because they have superior discrimination ability and a more conservative response criterion. However, even experts are influenced by irrelevant or misleading contextual information and often show evidence of confirmation bias. Worryingly, among forensic experts (including fingerprinting experts), only 52% regarded bias as a matter for concern and even fewer (26%) believed their own judgements were influenced by bias.
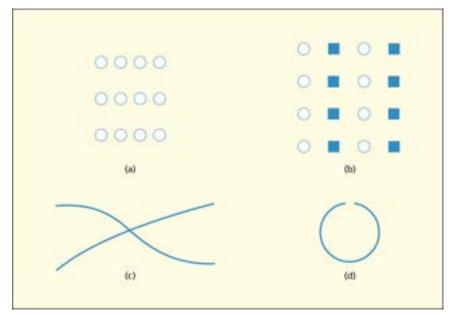
by the gestaltists, German psychologists (including Koffka, Köhler and Wertheimer) who emigrated to the United States between the two World Wars. Their fundamental principle was the **law of Prägnanz** – we typically perceive the simplest possible organisation of the visual field.

Most of the gestaltists' other laws can be subsumed under the law of Prägnanz. Figure 3.4(a) illustrates the law of proximity (visual elements close in space tend to be grouped together). Figure 3.4b shows the law of similarity (similar elements tend to be grouped together).

We see two crossing lines in Figure 3.4(c) because, according to the law of law continuation, we group together those elements requiring the fewest changes or interruptions in straight or smoothly curving lines. Finally, Figure 3.4(d) illustrates the law of closure: the missing parts of a figure are filled in to complete the figure (here a circle).

We might dismiss these principles as "mere textbook curiosities" (Wagemans et al., 2012a, p. 1180). However, the various grouping principles "pervade virtually all perceptual experiences because they determine the objects and parts that people perceive in their environment" (Wagemans et al., 2012a, p. 1180).

The gestaltists emphasised **figure-ground segmentation** in perception. The figure is perceived as having a distinct form or shape whereas the ground lacks form. In addition, the figure is perceived as being in

KEY TERMS

Law of Prägnanz
The notion that the simplest possible organisation of the visual environment is perceived; proposed by the gestaltists.

Figure-ground segmentation
The perceptual organisation of the visual field into a figure (object of central interest) and a ground (less important background).

Figure 3.4
Examples of the Gestalt laws of perceptual organisation: (a) the law of proximity; (b) the law of similarity; (c) the law of good continuation; and (d) the law of closure.

Figure 3.5
An ambiguous drawing that can be seen as either two faces or as a goblet.

front of the ground and the contour separating the figure from ground "belongs" to the figure. Check these claims with the faces-goblet illusion (see Figure 3.5). When the goblet is perceived as the figure, it seems to be in front of a dark background. Faces are in front of a light background when forming the figure.

What determines which region is identified as the figure and which as the ground? Regions that are convex (curving outwards), small, surrounded and symmetrical are most likely to be perceived as figures (Wagemans et al., 2012a). For example, Fowlkes et al. (2007) found with images of natural scenes that regions identified by observers as figures were generally smaller and more convex than ground regions.

Finally, the gestaltists argued perceptual grouping and organisation are innate or intrinsic to the brain. As a result, they de-emphasised the importance of past experience.

Findings

The gestaltists' approach was limited because they mostly used artificial figures, making it important to see whether their findings apply to more realistic stimuli. Geisler et al. (2001) used pictures to study the contours of flowers, rivers, trees and so on. They discovered object contours could be calculated accurately using two principles that were different from those emphasised by the gestaltists:

(1) Adjacent segments of any contour typically have very similar orientations.
(2) Segments of any contour that are further apart generally have somewhat different orientations.

Geisler et al. (2001) asked observers to decide which of two complex patterns presented together contained a winding contour. Task performance was well predicted by the two key principles described above.

Elder and Goldberg (2002) analysed the statistics of natural contours and obtained findings largely consistent with Gestalt laws. Proximity was a very powerful cue when deciding which contours belonged to which objects. There was also a small contribution from similarity and good continuation.

Numerous cues influence figure-ground segmentation and the perception of object boundaries with natural scenes. Mély et al. (2016) found colour and luminance (see Glossary) strongly influenced the perception of object boundaries. There was more accurate perception of object boundaries when several cues were combined than that found for any single cue in isolation.

In sum, there is some support for Gestalt laws in natural scene perception. However, figure-ground segmentation is more complex in natural scenes than most artificial figures, and so the Gestalt approach is oversimplified.

The gestaltists failed to discover several principles of perceptual organisation. For example, Palmer and Rock (1994) proposed the principle of uniform connectedness. According to this principle, any connected region having uniform visual properties (e.g., colour; texture; lightness) tends to be organised as a single perceptual unit. Palmer and Rock found grouping by uniform connectedness dominated proximity and similarity when there was a conflict.

Pinna et al. (2016) argued that the gestaltists de-emphasised the role of *dissimilarity* in perceptual organisation. Consider Figure 3.6. The perception of the empty circles as a rotated square or a diamond is strongly influenced by the location of the dissimilar element (i.e., the black circle). This illustrates the principle of *accentuation*: "Elements group in the same oriented direction of the dissimilar element placed . . . outside a whole set of continuous/homogeneous components" (Pinna et al., 2016, p. 21).

Much processing involved in perceptual organisation occurs very rapidly. Williford and von der Heydt (2016) discovered signals from neurons in V2 (see Chapter 2) relating to figure-ground organisation emerged within 70 ms of stimulus presentation for complex natural scenes as well as for simple figures. This extremely rapid processing is consistent with the gestaltists' assumption that perceptual organisation is due to innate factors but may also reflect massive experience in object recognition.

The role of learning was discussed by Bhatt and Quinn (2011). Infants as young as 3 or 4 months show grouping by continuation, proximity and connectedness, which is apparently consistent with the Gestalt position. However, other grouping principles (e.g., closure) were used only later in infancy, and infants typically made increased use of grouping principles over time. Thus, learning is important.
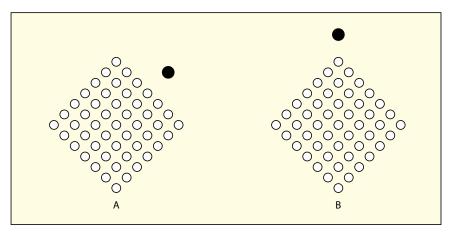
> **KEY TERM** 🔑
>
> **Uniform connectedness**
> The notion that adjacent regions in the visual environment having uniform visual properties (e.g., colour) are perceived as a single perceptual unit.

Figure 3.6
The dissimilar element (black circle) accentuates the tendency to perceive the array of empty circles as (A) a rotated square or (B) a diamond.
From Pinna et al., 2016.

According to the gestaltists, perceptual grouping occurs rapidly and should be uninfluenced by attentional processes. The evidence is mixed. Rashal et al. (2017) conducted several experiments. Attention was not required with grouping by proximity or similarity in colour. However, attention was required with grouping by similarity in shape. In general, attention was more likely to be required when the processes involved in perceptual grouping were relatively complex. Overall, the processes involved in perceptual grouping are much more complicated and variable than the gestaltists had assumed.

The gestaltists also assumed figure-ground segmentation is innate and so not reliant on past experience or learning. Barense et al. (2012) reported contrary evidence. Amnesic patients (having severe memory problems) and healthy controls were presented with various stimuli, some containing parts of well-known objects (see Figure 3.7). In other stimuli, the object parts were rearranged. The task was to decide which region of each stimulus was the figure.

The healthy controls identified the regions containing familiar objects as figures more often than those containing rearranged parts. In contrast, the amnesic patients showed no difference between the two types of stimuli because they experienced difficulty in identifying the objects presented. Thus, figure-ground segmentation can depend on past experience and memory (i.e., object familiarity).

Several recent theories explain perceptual grouping and figure-ground segmentation. For example, consider Froyen et al.'s (2015) Bayesian hierarchical grouping model, according to which observers initially form "beliefs" concerning the objects to be expected in the current context. In addition, their visual system assumes the visual image consists of a mixture of objects. The information available in the image is then used to change the subjective probabilities of different grouping hypotheses to make optimal use of that information. Of key importance, observers use their learned knowledge of patterns and objects (e.g., visual elements close together generally belong to the same object).

The above approach exemplifies theories based on Bayesian inference (see Glossary). Their central assumption is that the initial subjective probabilities associated with various hypotheses as to the organisation of objects within a visual image *change* on the basis of the information it provides. This approach is much more realistic than the gestaltists' relatively cut-and-dried approach.
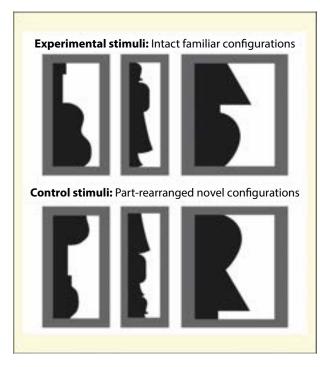

Figure 3.7
The top row shows intact familiar shapes (from left to right: a guitar, a standing woman, a table lamp). The bottom row shows the same objects but with the parts rearranged. The task was to decide which region in each stimulus was the figure.

From Barense et al. (2012). Reprinted with permission of Oxford University Press.

Evaluation

What are the strengths of the Gestalt approach? First, the gestaltists focused on key issues (e.g., figure-ground segmentation). Second, nearly all their grouping laws (and the notion of figure-ground segmentation) have stood the test of time and are applicable to natural scenes as well as artificial figures. Third, the notion that observers perceive the simplest possible organisation of the visual environment has proved very fruitful. Many recent theories are based on the assumption that striving for simplicity is central to visual perception (Jäkel et al., 2016).

What are the approach's limitations? First, the gestaltists de-emphasised the importance of past experience and learning. As Wagemans et al. (2012b, p. 1229) pointed out, the gestaltists "focused almost exclusively on processes intrinsic to the perceiving organism . . . The environment itself did not interest [them]".

Second, the gestaltists produced *descriptions* of important perceptual phenomena but not adequate *explanations*. Recently, however, such explanations have been provided.

Third, nearly all the evidence the gestaltists provided was based on two-dimensional drawings. The greater complexity of real-world scenes (e.g., important parts of objects hidden or occluded) means additional explanatory assumptions are required.

Fourth, the gestaltists did not discover all the principles of perceptual organisation. Among such undiscovered principles are uniform connectedness, the principle of accentuation and generalised common fate (e.g., when elements of a visual scene become brighter or darker together, they are grouped together). More generally, the gestaltists did not appreciate the sheer complexity of the processes involved in perceptual grouping.

Fifth, the gestaltists focused mostly on drawings involving only *one* Gestalt law. With natural scenes, several laws often operate simultaneously and interact in complex ways not predicted by the gestaltists (Jäkel et al., 2016).

Sixth, the gestaltists' approach was too *inflexible*. They did not realise perceptual grouping and figure-ground segregation depend on complex interactions between basic (and possibly innate) processes and past experience (Rashal et al., 2017).

APPROACHES TO OBJECT RECOGNITION

Object recognition (identifying objects in the visual field) is enormously important if we are to interact effectively with the environment. We start with basic aspects of the human visual system followed by major theories of object recognition.

Perception-action model

Milner and Goodale's (1995, 2008) perception-action model (discussed in Chapter 2) is relevant to understanding object perception. It is based on a distinction between ventral (or "what") and dorsal (or "how") streams (see

Figure 2.9), with the latter providing visual guidance for action (e.g., grasping). They argued object recognition and perception depend primarily on the ventral stream. This stream is hierarchically organised. Visual processing basically proceeds from the retina through several areas including the lateral geniculate nucleus, V1, V2 and V4, culminating in the inferotemporal cortex.

The importance of the ventral stream is indicated by research showing object recognition can be reasonably intact after damage to the dorsal stream (Goodale & Milner, 2018). However, object recognition involves numerous *interactions* between the ventral and dorsal streams (Freud et al., 2017b).

Spatial frequency

Visual perception develops over time even though it seems instantaneous (Hegdé, 2008). The visual processing involved in object recognition typically proceeds in a coarse-to-fine way with initial coarse or general processing followed by fine or detailed processing. As a result, we can perceive visual scenes at a very general level and/or at a fine-grained level.

How does coarse-to-fine processing occur? Numerous cells in the primary visual cortex respond to high spatial frequencies and capture fine detail in the visual image. Numerous others respond to low spatial frequencies and capture coarse information in the visual image.

Low spatial frequency information (often relating to motion and/or spatial location) is transmitted rapidly to higher-order brain areas via the fast magnocellular system using the dorsal visual stream (discussed in Chapter 2). Awasthi et al. (2016) used red light to produce magnocellular suppression. As predicted, this interfered with the low spatial frequency components of face perception.

In contrast, high spatial frequency information (often relating to colour, shape and other aspects of object recognition) is transmitted relatively slowly via the parvocellular system using the ventral visual stream (see Chapter 2). This speed difference explains why coarse processing typically precedes fine processing, although conscious perception is typically based on integrated low and high spatial information.

We can observe the effects of varying spatial frequency by comparing images consisting only of low or high spatial frequency (see Figure 3.8). You probably agree it is considerably easier to achieve object recognition with the high spatial frequency image.

Figure 3.8
High and low spatial frequency versions of a place (a building).
From Awasthi et al. (2016).

Findings

Musel et al. (2012) presented participants with very brief (150 ms) scenes proceeding from coarse (low spatial frequency) to fine (high spatial frequency) or vice versa (sample videos can be viewed at DOI.10.1371/journal.pone.003893). Performance (deciding whether each scene was an outdoor or indoor
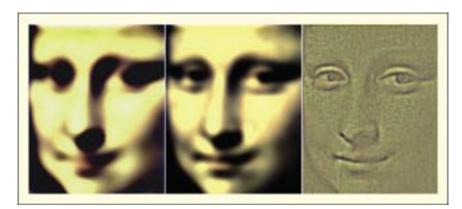
Figure 3.9
Image of *Mona Lisa* revealing very low spatial frequencies (left), low spatial frequencies (centre) and high spatial frequencies (right).

From Livingstone (2000). By kind permission of Margaret Livingstone.

one) was faster with the coarse-to-fine sequence, a finding subsequently replicated by Kauffmann et al. (2015). These findings suggest the visual processing of natural scenes is predominantly coarse to fine. This sequence may be more effective because low spatial frequency information is used to generate plausible interpretations of the visual input.

There is considerable evidence that global (general) processing often precedes local (specific) processing (discussed on p. 95). Much research has established an association between processing low spatial frequencies and global perception and between processing high spatial frequencies and local perception. However, use of low and high spatial frequency information in visual processing is often very *flexible* and is influenced by task demands. Flevaris and Robertson (2016, p. 192) reviewed research showing, "Attention to global and local aspects of a display biases the flexible selection of *relatively* lower and *relatively* higher SFs [spatial frequencies] during image processing."

Finally, we can explain the notoriously elusive smile of Leonardo da Vinci's *Mona Lisa* with reference to spatial frequencies. Livingston (2000) produced images of that painting with different spatial frequencies. Mona Lisa's smile is much more obvious in the two low spatial frequency images (see Figure 3.9). Livingston pointed out that our central or foveal vision is dominated by higher spatial frequencies compared with our peripheral vision. As a result, "You can't catch her smile by looking at her mouth. She smiles until you look at her mouth" (p. 1299).

Historical background: Marr's computational approach

David Marr (1982) proposed a very influential theory. He argued object recognition involves various processing stages and is much more complex than had previously been thought. More specifically, Marr claimed observers construct various representations (descriptions) providing increasingly detailed information about the visual environment:

- *Primal sketch*: this provides a two-dimensional description of the main light intensity changes in the visual input, including information about edges and contours.

- *2½-D sketch*: this incorporates a description of the depth and orientation of visible surfaces using information from shading, texture, motion and binocular disparity. It resembles the primal sketch in being viewer-centred or viewpoint-dependent (i.e., it is influenced by the angle from which the observer sees objects or the environment).
- *3-D model representation*: this describes objects' shapes and their relative positions three-dimensionally; it is *independent* of the observer's viewpoint and so is viewpoint-invariant.

Why has Marr's theoretical approach been so influential? First, he successfully combined ideas from neurophysiology, anatomy and computer vision (Mather, 2015). Second, he was among the first to realise the enormous complexity of object recognition. Third, his distinction between viewpoint-dependent and viewpoint-invariant representations triggered much subsequent research (discussed on pp. 109–111).

What are the limitations of Marr's approach? First, he focused excessively on bottom-up processes. Marr (1982, p. 101) admitted, "Top-down processing is sometimes used and necessary." However, he de-emphasised the major role expectations and knowledge play in object recognition (discussed in detail on pp. 111–116).

Second, Marr assumed that "Vision tells the truth about what is out there" (Mather, 2015, p. 44). In fact, there are numerous exceptions. For example, people observed from a tall building (e.g., the Eiffel Tower) seem very small. Another example is the vertical-horizontal illusion – observers typically overestimate the length of a vertical line when it is compared against a horizontal line of the same length (e.g., Gavilán et al., 2017).

Third, many processes proposed by Marr are incredibly complex computationally. As Mather (2015, p. 44) pointed out, "The computations required to produce view-independent 3-D object models are now thought by many researchers to be too complex."

Biederman's recognition-by-components theory

Biederman's (1987) recognition-by-components theory developed Marr's theoretical approach. His central assumption was that objects consist of basic shapes or components known as "geons" (geometric ions); examples include blocks, cylinders, spheres, arcs and wedges. Biederman claimed there are approximately 36 different geons, which sounds suspiciously low to provide descriptions of all objects. However, geons can be combined in almost endless ways. For example, a *cup* is an arc connected to the side of a cylinder. A *pail* involves the same two geons but with the arc connected to the top of the cylinder.

Figure 3.10 shows the key features of recognition-by-components theory. We have already considered the stage where the components or geons of an object are determined. When this information is available, it is matched with stored object representations or structural models consisting of information about the nature of the relevant geons, their orientations, sizes and so on. Whichever stored representation fits best with the geon-based information obtained from the visual object determines which object is identified by observers.

As indicated in Figure 3.10, the first step in object recognition is edge extraction in which various aspects of the visual stimulus (e.g., luminance; texture; colour) are processed, leading to a description of the object resembling a line drawing. After that, decisions are made as to how the object should be segmented to establish its geons.

Which edge information should observers focus on? According to Biederman (1987), non-accidental image properties are crucial. These are aspects of the visual image that are invariant across different viewing angles. Examples include whether an edge is straight or curved and whether a contour is concave (hollow) or convex (bulging) with the former of particular importance. Biederman assumed objects' geons of a visual object are constructed from various non-accidental or invariant properties.

This part of the theory leads to the key prediction that object recognition is typically viewpoint-invariant (i.e., objects can be recognised equally easily from nearly all viewing angles). The argument is that object recognition depends crucially on the identification of geons, which can be identified from numerous viewpoints. Thus, object recognition is difficult only when one or more geons are hidden from view.
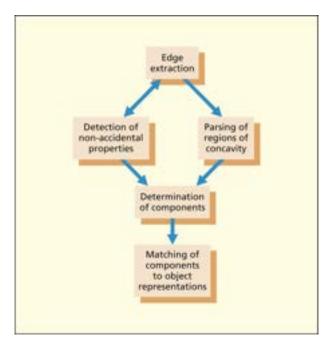
How do we recognise objects in suboptimal viewing conditions (e.g., an intervening object obscures part of the target object)? First, non-accidental properties can still be detected even when only parts of edges are visible. Second, if the *concavities* of a contour are visible, there are mechanisms for restoring the missing parts of the contour. Third, we can recognise many objects when some geons are missing because there is much redundant information under optimal viewing conditions.

Findings

Non-accidental properties play a vital role in object recognition (Parker & Serre, 2015). For example, it is easier to distinguish between two objects differing in non-accidental properties. In addition, neuroimaging studies

Irving Biederman. University of Southern California.

Figure 3.10

An outline of Biederman's recognition-by-components theory.

Adapted from Biederman (1987).

reveal greater neural responses to changes in non-accidental properties than other visual changes. Rolls and Mills (2018) developed a model of object recognition showing how non-accidental properties of objects can promote viewpoint-invariant object recognition.

There is general agreement that an object's contour or outline is important in object recognition. For example, camouflage in many animal species is achieved by markings breaking up and distorting contour information (Webster, 2015). There is also general agreement that concavities and convexities are especially informative regions of an object's contour. However, the evidence relating to Biederman's (1987) assumption that concavity is more important than convexity in object recognition is mixed (Schmidtmann et al., 2015).

In their own study, Schmidtmann et al. (2015) focused specifically on shape recognition using unfamiliar shapes. Shape recognition depended more on information about convexities than concavities (although concavity information had some value). They argued convexity information is likely to be more important because convexities reveal an object's outer boundary.

According to the theory, object recognition depends on edge rather than surface information (e.g., colour). However, Sanocki et al. (1998) argued that edge-extraction processes are less likely to produce accurate object recognition when objects are presented in the context of other objects rather than on their own. This is because it can be hard to decide which edges belong to which objects when several objects are presented together. Sanocki et al. presented observers briefly with objects, with line drawings, or full-colour photographs of objects. As predicted, object recognition was much worse with the line drawings than the full-colour photographs when objects were presented in context.

A key theoretical prediction is that object recognition is typically viewpoint-invariant. Biederman and Gerhardstein (1993) supported this prediction when familiar objects presented at different angles were named rapidly. However, numerous other studies have failed to obtain evidence for viewpoint-invariance. This is especially the case with unfamiliar objects differing from familiar objects in not having previously been viewed from multiple viewpoints (discussed in next section on pp. 109–111).

Evaluation

Biederman's (1987) recognition-by-components theory has been very influential. It indicates how we can identify objects despite substantial differences among the members of most categories in shape, size and orientation. The assumption that non-accidental properties of stimuli and geons play a role in object recognition has received much support.

What are the theory's limitations? First, it focuses predominantly on bottom-up processes triggered directly by the stimulus input. As a result, it de-emphasises the impact on object recognition of top-down processes based on expectation (Trapp & Bar, 2015; discussed further on pp. 111–116). Second, the theory accounts only for fairly unsubtle perceptual discriminations. It cannot explain how we decide whether an animal is, for example, a particular breed of dog or cat. Third, the notion that

objects consist of invariant geons is too *inflexible*. As Hayward and Tarr (2005, p. 67) pointed out, "You can take almost any object, put a working light-bulb on the top, and call it a lamp."

Does viewpoint influence object recognition?

Form a visual image of a bicycle. Your image probably involved a side view with both wheels clearly visible. We can use this example to discuss a theoretical controversy. Consider an experiment where some participants see a photograph of a bicycle in the typical (or canonical) view as in your visual image, whereas others see a photograph of the same bicycle viewed end-on or from above. Would those given the typical view identify the object as a bicycle fastest?

We will address the above question shortly. Before that, we must discuss two key terms mentioned earlier. If object recognition is equally rapid and easy regardless of viewing angle, it is *viewpoint-invariant*. In contrast, if object recognition is faster and easier when objects are seen from certain angles, it is *viewer-centred* or *viewpoint-dependent*. Another important distinction is between *categorisation* (e.g., is the object a dog?) and *identification* (e.g., is the object a poodle?), which requires within-category discriminations.

Findings

Milivojevic (2012) reviewed behavioural research in this area. Object recognition is typically uninfluenced by an object's orientation when categorisation is required (i.e., it is viewpoint-invariant). In contrast, object recognition is significantly slower if an object's orientation differs from its canonical or typical viewpoint when identification is required (i.e., it is viewer-centred). Hamm and McMullen (1998) reported supporting findings. Changes in viewpoint had no effect on speed of object recognition when categorisation was required (e.g., deciding an object was a car). However, there were clear effects of changing viewpoint with identification (e.g., deciding whether an object was a taxi).

Small (or non-significant) effects of object orientation on categorisation time do *not* necessarily indicate orientation has not affected internal processing. Milivojevic et al. (2011) found stimulus orientation had only small effects on speed and accuracy of categorisation. However, early components of the event-related potentials (ERPs; see Glossary) were larger when stimuli were not in the upright position. Thus, stimulus orientation had only modest effects on task performance but perceptual processing was less demanding with upright stimuli.

Neuroimaging research has enhanced our understanding of object recognition (Milivojevic, 2012). With categorisation tasks, brain activation is mostly very similar regardless of object orientation. However, orientation influences brain activity *early* in processing suggesting initial processing is viewpoint-dependent.

With identification tasks, there is typically greater activation of areas within the inferior temporal cortex when objects are not in their typical or canonical orientation (Milivojevic, 2012). This finding is unsurprising

since the inferotemporal cortex is heavily involved in object recognition (Gauthier & Tarr, 2016). Identification may require additional processing (e.g., more detailed processing of object features) for objects presented in unusual orientations.

Learning influences the extent to which object recognition is viewpoint-dependent or viewpoint-invariant. Zimmermann and Eimer (2013) presented unfamiliar faces on 640 trials. Face recognition was viewpoint-dependent initially but became more viewpoint-invariant there-after. Learning caused more information about each face to be stored in long-term memory and this facilitated rapid access to visual face memory regardless of facial orientation.

Etchells et al. (2017) also studied the effects of learning on face recognition. During learning, observers were repeatedly shown one or two views of unfamiliar faces. Subsequently they were shown a novel view of these faces. There was evidence of viewpoint-invariant face recognition when learning had been based on two different views but *not* when it had been based on only a single view.

Related research was reported by Weibert et al. (2016). They found evidence of a viewpoint-invariant response in face-selective regions of the medial temporal lobe with familiar (but not unfamiliar) faces. Thus, viewpoint-invariant responses during object recognition are more frequent for faces for which observers have stored considerable relevant information.

Evidence of viewpoint-dependent or viewpoint-invariant responses within the brain often depends on the precise brain areas studied. Erez et al. (2016) found viewpoint-dependent responses in several visual areas (e.g., fusiform face area) but viewpoint-invariant responses in the perirhinal cortex. There is more evidence for viewpoint-invariant brain responses *late* rather than *early* in visual processing. Why is that? As Erez et al. (p. 2271) argued, "Representations of low-level features are transformed into more complex and invariant representations as information flows through successive stages of [processing]."

Most research is limited because object recognition is typically assessed in only *one* context, which may prompt either viewpoint-invariant or viewpoint-dependent recognition performance. Tarr and Hayward (2017) argued this approach can misleadingly suggest observers store *only* viewpoint-invariant or viewpoint-dependent information. Accordingly, they used various contexts. Observers originally learned the identities of novel objects that could be discriminated by viewpoint-invariant information. As predicted, they exhibited viewpoint-invariant object recognition when tested. When the testing context was changed to make it hard to continue to use that approach, observers shifted to exhibiting viewpoint-dependent behaviour.

The central conclusion from the above findings is that: "Object representations are neither viewpoint-dependent nor viewpoint-invariant, but rather encode multiple kinds of information . . . deployed in a flexible manner appropriate to context and task" (Tarr & Hayward, 2017, p. 108). Thus, visual object representations contain richer and more variegated information than typically assumed on the basis of limited testing conditions.

Conclusions

As Gauthier and Tarr (2016, p. 179) concluded: "Depending on the exper-imental conditions and which parts of the brain we look at, one can obtain data supporting both the structural-description (i.e., the viewpoint-invariant) and the view-based [viewpoint-dependent] approaches." There has been progress in identifying factors (e.g., is categorisation or identifi-cation required? is the object familiar or unfamiliar?) influencing whether object recognition is viewpoint-invariant or viewpoint-dependent.

Gauthier and Tarr (2016, p. 379) argued researchers should address the following question: "What is the nature of the features that com-prise high-level visual representations and lead to image-dependence or image-invariance?" Thus, we should focus more on *why* object recognition is viewpoint-invariant or viewpoint-dependent. As yet, "The exact and fine-grained features of object representations are still unknown and are not easily resolved" (Gauthier & Tarr, 2016, p. 379).

OBJECT RECOGNITION: TOP-DOWN PROCESSES

Historically, most theorists (e.g., Marr, 1982; Biederman, 1987) studying object recognition emphasised bottom-up processes. Apparent support can be found in the hierarchical nature of visual processing. As Yardley et al. (2012, p. 4) pointed out,

> Traditionally, visual object recognition has been taken as mediated by a hierarchical, bottom-up stream that processes an image by systematically analysing its individual elements and relaying this information to the next areas until the overall form and identity are determined.

The above account, assuming a feedforward hierarchy of processing stages from visual cortex through to inferotemporal cortex, is oversimplified. There are as many *backward* projecting neurons (associated with top-down processing) as *forward* projecting ones throughout most of the visual system (Gilbert & Li, 2013). Up to 90% of the synapses from incoming neurons to primary visual cortex (involved in early visual processing) originate in the cortex and thus reflect top-down processes. Recurrent processing (a form of top-down processing) from higher to lower brain areas is often necessary for conscious visual perception (van Gaal & Lamme, 2012; see Chapter 16).

Top-down processes should have their greatest impact on object rec-ognition when bottom-up processes are relatively *uninformative* (e.g., when observers are presented with degraded or briefly presented stimuli). Support for this prediction is discussed on p. 112.

Findings

Evidence for the involvement of top-down processes in visual percep-tion was reported by Goolkasian and Woodberry (2010). They presented observers with *ambiguous* figures immediately preceded by primes relevant to one interpretation (see Figure 3.11). The primes systematically biased the interpretation of the ambiguous figures via top-down processes.
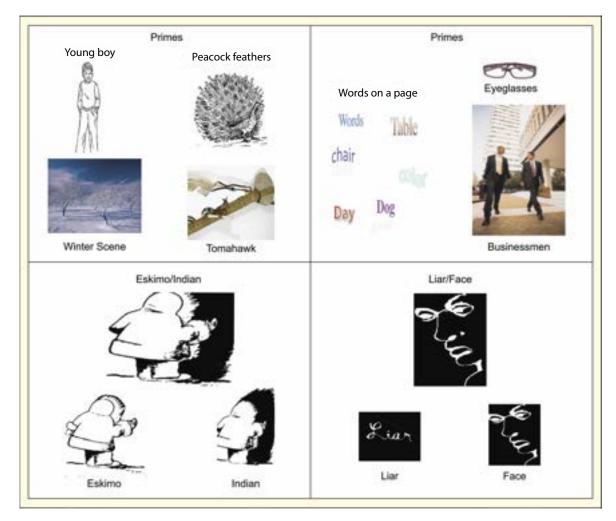
Figure 3.11

Ambiguous figures (e.g., Eskimo/Indian, Liar/Face) were preceded by primes (e.g., Winter Scene, Tomahawk) relevant to one interpretation of the following figure.

From Goolkasian and Woodberry (2010). Reprinted with permission from the Psychonomic Society 2010.

Viggiano et al. (2008) obtained strong evidence that top-down processes within the prefrontal cortex influence object recognition. Observers viewed blurred or non-blurred photographs of living and non-living objects. On some trials, repetitive transcranial magnetic stimulation (rTMS; see Glossary) was applied to the dorsolateral prefrontal cortex to disrupt top-down processing. rTMS slowed object recognition time only with blurred photographs. Thus, top-down processes were directly involved in object recognition when the sensory information available to bottom-up processes was limited.

Controversy

Firestone and Scholl (2016) argued in a review, "There is . . . no evidence for top-down effects of cognition on visual perception." They claimed that top-down processes often influence response bias, attention or memory rather than perception itself.

First, consider ambiguous or reversible figures (e.g., the faces-goblet illusion shown in Figure 3.5). Observers alternate between the two possible interpretations (e.g., faces vs goblet). The dominant one at any moment depends on their direction of attention but not necessarily on top-down processes (Long & Toppino, 2004).

Second, Auckland et al. (2007) presented observers briefly with a target object (e.g., *playing cards*) surrounded by four context objects. When the context objects were semantically related to the target (e.g., *dice*; *chess pieces*; *plastic chips*; *dominoes*), the target was recognised more often than when they were semantically unrelated. This finding depended in part on response bias (i.e., guesses based on context) rather than perceptual information about the target. (More evidence of response bias is discussed in the Box).

Firestone and Scholl's (2016) article has provoked much controversy. Lupyan (2016, p. 40) attacked their tendency to attribute apparent top-down effects on perception to attention, memory and so on: "This 'It's not perception, it's just X' reasoning assumes that attention, memory, and so forth be cleanly split from *perception proper*." In fact, all these processes interact dynamically and so attention, perception and memory are not clearly separate.

IN THE REAL WORLD: SHOOTER BIAS

Shooter bias is shown by "More shooting errors for unarmed black than white suspects" (Cox & Devine, 2016, p. 237). Black Americans are more than twice as likely as white Americans to be unarmed when killed by the police (Ross, 2015). For example, on 22 November 2014, a police officer in Cleveland, Ohio, shot dead a 12-year-old black male (Tamir Rice) playing with a replica pistol.

Shooter bias may reflect top-down influences on visual perception. Payne (2006) presented a white or black face followed by the very brief presentation of a gun or tool. When participants made a rapid response, they indicated falsely they had seen a gun more often when the face was black.

Shooter bias reflects top-down effects based on inaccurate racial stereotypes associating black individuals with threat (e.g., Azevedo et al., 2017). This bias might be due to *direct* top-down effects on perception: objects are more likely to be *misperceived* as guns if held by black individuals. Alternatively, shooter bias may reflect *response bias* (the expectation someone has a gun is greater if that person is black rather than white): there is no effect on perception but shooters require less perceptual evidence to shoot a black individual.

Azevedo et al. (2017) found a briefly presented weapon (a gun) was more accurately perceived when preceded by a black face than a white one. However, the opposite was the case when a tool was presented. These findings were due to response bias rather than perception.

Moore-Berg et al. (2017) asked non-black participants to decide rapidly whether or not to shoot an armed or unarmed white or black person of high or low socio-economic status. There was shooter bias: participants were biased towards shooting if the individual was black, of low socio-economic status, or both. This shooter bias mostly reflected a response bias against shooting a white person of high socio-economic status (probably because of a low level of perceived danger).

Figure 3.12
A brick wall that can be seen as something else.
From Plait (2016).

Further findings

Howe and Carter (2016) identified two perception-like phenomena driven by top-down processes. First, there are visual hallucinations which are found in schizophrenic patients. Hallucinations are experienced as actual perceptions even though the relevant object is not present and so they cannot depend on bottom-up processes. Second, there is visual imagery (discussed on pp. 130–137). Visual imagery involves several processes involved in visual perception. Like hallucinations, visual imagery occurs in the absence of bottom-up processes because the relevant object is absent.

Lupyan (2017) discussed numerous top-down effects on visual perception in studies avoiding the problems identified by Firestone and Scholl (2016). Look at Figure 3.12, which apparently shows an ordinary brick wall. If that is what you see, have another look. This time see whether you can spot the object mentioned at the end of the Conclusions section. Once you have spotted the object, it becomes impossible *not* to see it afterwards.

Here we have powerful effects of top-down processing on perception based on knowledge of what is in the photograph.

Conclusions

As Firestone and Scholl (2016) argued, it is hard to demonstrate top-down processes *directly* influence perception rather than attention, memory or response bias. However, many studies have shown such a direct influence. As Yardley et al. (2012, p. 1) pointed out, "Perception relies on existing knowledge as much as it does on incoming information." Note, however, the influence of top-down processes is generally greater when visual stimuli are degraded. By the way, the hard-to-spot object in the photograph is a cigar!

Theories emphasising top-down processes

Bar et al. (2006) found greater activation of the orbitofrontal cortex (part of the prefrontal cortex) when object recognition was hard than when it was easy. This activation occurred 50 ms *before* activation in recognition-related regions of the temporal cortex, and so seemed important for object recognition. In Bar et al.'s model, object recognition depends on top-down processes involving the orbitofrontal cortex and bottom-up processes involving the ventral visual stream (see Figure 3.13; and Chapter 2).

Trapp and Bar (2015) developed this model claiming that visual input rapidly elicits various hypotheses concerning what has been presented. Subsequent top-down processes associated with the orbitofrontal cortex select relevant hypotheses and suppress irrelevant ones. More specifically, the orbitofrontal cortex uses contextual information to generate hypotheses and resolve competition among hypotheses. Palmer (1975) showed the importance of context. He presented a picture of a scene (e.g., a *kitchen*) followed by the very brief presentation of the picture of an object. The object was recognised more often when relevant to the context (e.g., a *loaf*) than when irrelevant (e.g., a *drum*).

Interactive-iterative framework

Baruch et al. (2018) argued that previous theorists had not appreciated the full complexities of interactions between bottom-up and top-down processes in object recognition. They rectified this situation with their interactive-iterative framework (see Figure 3.14).

According to this framework, observers typically form hypotheses concerning object identity based on their goals, knowledge and the environmental context. Of importance, these hypotheses are often formed *before* the object is presented. Observers discriminate among competing hypotheses by attending to a distinguishing feature of the object. For example, if your tentative hypothesis was *elephant*, you might allocate attention to the expected location of its trunk. If that failed to provide the necessary information because that area was partially hidden (see Figure 3.15), you might then attend to other features (e.g., size and shape of the leg; skin texture).

In sum, Baruch et al. (2018) emphasised two related top-down processes strongly influencing object recognition. First, observers form hypotheses about the possible identity of an object *prior to* (or in interaction with) the visual input. Second, observers direct their attention to object parts likely to be maximally informative concerning its identity.
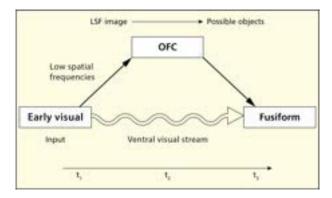

Figure 3.13

In this modified version of Bar et al.'s (2006) theory, it is assumed that object recognition involves two different routes: (1) a top-down route in which information proceeds rapidly to the orbitofrontal cortex, which is involved in generating predictions about the object's identity; (2) a bottom-up route using the slower ventral visual stream.

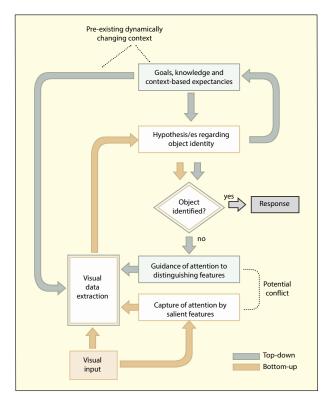From Yardley et al. (2012). Reprinted with permission from Springer.

Figure 3.14

Interactive-iterative framework for object recognition with top-down processes shown in dark green and bottom-up processes in brown.

From Baruch et al. (2018). Reprinted with permission of Elsevier.

Figure 3.15
Recognising an elephant when a key feature (its trunk) is partially hidden.
From Baruch et al. (2018). Reprinted with permission of Elsevier.

Findings

According to the interactive-iterative framework, expectations can exert top-down influences on processing even *before* a visual stimulus is presented. Kok et al. (2017) obtained support for this prediction. Observers expecting a given stimulus produced a neural signal resembling that generated by the actual presentation of the stimulus shortly before it was presented.

Baruch et al. (2018) tested various predictions from their theoretical framework. In one experiment, participants decided which of two types of artificial fish (tass or grout) had been presented. The two fish types differed with respect to distinguishing features associated with the tail and the mouth, with the tail being easier to discriminate. As predicted, participants generally attended more to the tail than the mouth region from stimulus onset. When much of the tail region was hidden from view, participants redirected their attention to the mouth region.

Summary

Numerous theorists have argued that object recognition depends on top-down processes as well as bottom-up ones. Baruch et al.'s (2018) interactive-iterative framework extends such ideas by identifying how these two types of processes interact. Of central importance, top-down processes influence the allocation of attention, and the allocation of attention influences subsequent bottom-up processing.

FACE RECOGNITION

There are two main reasons for devoting a section to face recognition. First, recognising faces is of enormous importance to us, since we generally identify individuals from their faces. Form a visual image of someone important in your life – it probably contains detailed information about their face.

Second, face recognition differs importantly from other forms of object recognition. As a result, we need theories specifically devoted to face recognition rather than simply relying on theories of object recognition.

Face vs object recognition

How does face recognition differ from object recognition? There is more holistic processing in face recognition. **Holistic processing** involves "integration across the area of the face, or processing of the relationships between features as well as, or instead of, the features themselves" (Watson & Robbins, 2014, p. 1). Holistic processing is *faster* because facial features

KEY TERM 🗝

Holistic processing
Processing that involves *integrating* information from an entire object (especially faces).

are processed in parallel rather than individually. Caharel et al. (2014) found faces can be categorised as familiar or unfamiliar within approximately 200 ms. Holistic processing is also more *reliable* than feature processing because individual facial features (e.g., mouth shape) are subject to change.

Relevant evidence comes from the **face inversion effect**: faces are much harder to identify when presented inverted or upside-down rather than upright (Bruyer, 2011). This effect probably reflects difficulties in processing inverted faces holistically. There are surprisingly large effects of face inversion within the brain – Rosenthal et al. (2017, p. 4823) found face inversion "induces a dramatic functional reorganisation across related brain networks".

In contrast, adverse effects of inversion are often much smaller with non-face objects. For example, Klargaard et al. (2018) found there was a larger inversion effect for faces than for cars. However, it can be argued we possess expertise in face recognition and so we should consider individuals possessing expertise with non-face objects. The findings are mixed. Rossion and Curran (2010) found car experts had a much smaller inversion effect for cars than faces. However, those with the greatest expertise showed a greater inversion effect for cars. In contrast, Weiss et al. (2016) found horse experts had no inversion effect for horses.

More evidence suggesting faces are special comes from the **part-whole effect** – it is easier to recognise a face part when presented within a whole face rather than in isolation. Farah (1994) studied this effect using drawings of faces and houses. Participants' ability to recognise face parts was much better when whole faces were presented rather than only a single feature (i.e., the part-whole effect). In contrast, recognition performance for house features was very similar in whole and single-feature conditions.

Richler et al. (2011) explored the hypothesis that faces are processed holistically by using composite faces. Composite faces consist of a top half and a bottom half that may or may not be from the same face. The task was to decide whether the top halves of two successive composite faces were the same or different. Performance was worse when the bottom halves of the two composite faces were different. This composite face effect suggests people find it hard to ignore the bottom halves and thus that face processing is holistic.

Finally, accurate face recognition is so important to humans we might expect to find holistic processing of faces even in young children. As predicted, children aged between 3 and 5 show holistic processing (McKone et al., 2012).

In sum, face recognition (even in young children) involves holistic processing. However, it remains unclear whether the processing differences between faces and other objects occur because faces are special or because we have dramatically more expertise with faces than most other object categories. Relevant evidence was reported by Ross et al. (2018). When participants were presented with car pictures, car experts formed more holistic representations within the brain than did car novices. The role played by expertise is discussed further shortly.

> **KEY TERM** 🔑
>
> **Face inversion effect**
> The finding that faces are much harder to recognise when presented upside down; the effect of inversion is less marked (or absent) with other objects.
>
> **Part-whole effect**
> The finding that a face part is recognised more easily when presented in the context of a whole face rather than on its own.

KEY TERM 🔑

Prosopagnosia
A condition (also known as face blindness) in which there is a severe impairment in face recognition but much less impairment of object recognition; it is often the result of brain damage (acquired prosopagnosia) but can also be due to impaired development of face-recognition mechanisms (developmental prosopagnosia).

Prosopagnosia

Much research has involved brain-damaged patients with severely impaired face processing. Such patients suffer from **prosopagnosia** (pros-uh-pag-NO-see-uh) coming from the Greek words for "face" and "without knowledge". Prosopagnosia is also known as "face blindness".

Prosopagnosia is a heterogeneous or diverse condition with the precise problems of face and object recognition varying across patients. It can be caused by brain damage (acquired prosopagnosia) or can occur in the absence of any obvious brain damage (developmental prosopagnosia). Acquired prosopagnosics differ in terms of their specific face-processing deficits and brain areas involved (discussed later).

Studying prosopagnosics is of direct relevance to the issue of whether face recognition involves specific or specialised processes absent from object recognition. If prosopagnosics invariably have great impairments in object recognition, it would suggest face and object recognition involve similar processes. In contrast, if some prosopagnosics have intact object recognition, it would imply the processes underlying the two forms of recognition are different.

Farah (1991) reviewed research on patients with acquired prosopagnosia. All these patients also had more general problems with object recognition. However, some exceptions have been reported. Moscovitch

IN REAL LIFE: HEATHER SELLERS 🔍

We can understand the profound problems prosopagnosics suffer in everyday life by considering Heather Sellers (see YouTube: "You Don't Look Like Anyone I Know"). She is an American woman with severe prosopagnosia. When she was a child, she became separated from her mother at a grocery store. When reunited with her mother, she did not initially recognise her.

Heather Sellers still has difficulty in recognising her own face. Heather: "A few times I have been in a crowded elevator with mirrors all found and a woman will move, and I will go to get out the way and then realise 'oh that woman is me'." Such experiences made her very anxious.

Surprisingly, Heather Sellers was 36 before she realised she had prosopagnosia. Why was this? As a child, she became very skilled at identifying people by their hair style, body type, clothing, voice and gait. In spite of these skills, she has occasionally failed to recognise her own husband! According to Heather Sellers, "Not being able to reliably know who people are – it feels terrible like failing all the time."

Heather Sellers.
Patricia Roehling.

et al. (1997) studied CK, a man with object agnosia (impaired object recognition). He performed comparably to healthy controls on several face-recognition tasks including photos, caricatures and cartoons.

Geskin and Behrmann (2018) reviewed the literature on patients with developmental prosopagnosia. Out of 238 cases, 80% had impaired object recognition but 20% did not. Thus, several patients had impaired face recognition but not object recognition. We would have a double dissociation (see Glossary) if we could find individuals with developmental object agnosia but intact face recognition. Germine et al. (2011) found a female (AW), who had preserved face recognition but impaired object recognition for many categories of objects.

Overall, far more individuals have impaired face recognition (prosopagnosia) but relatively intact object recognition than have impaired object recognition but intact face recognition. These findings suggest that, "Face recognition is an especially difficult instance of object recognition where both systems [i.e., face and object recognition] rely on a common mechanism" (Geskin & Behrmann, 2018, p. 18). Face recognition is hard in part because it involves distinguishing among broadly similar category members (e.g., *two eyes*; *nose*; *mouth*). In contrast, object recognition often only involves identifying the relevant category (e.g., *cat*; *car*). According to this viewpoint, prosopagnosics would perform poorly if required to make fine-grained perceptual judgments with objects.

An alternative interpretation emphasises *expertise* (Wang et al., 2016). Nearly everyone has more experience (and expertise) at recognising faces than the great majority of other objects. It is thus possible that brain damage in prosopagnosics affects areas associated with expertise generally rather than specifically faces (the expertise hypothesis is discussed on pp. 122–124).
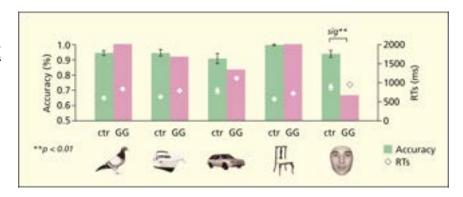
Findings

In spite of their poor conscious or explicit recognition of faces, many prosopagnosics show evidence of covert recognition (face processing without conscious awareness). For example, Eimer et al. (2012) found developmental prosopagnosics were much worse than healthy controls at explicit recognition of famous faces (27% vs 82% correct, respectively). However, famous faces produced brain activity in half the developmental prosopagnosics indicating the relevant memory traces were activated (covert recognition). These prosopagnosics have very poor explicit recognition performance because brain areas containing more detailed information about the famous individuals were not activated.

Busigny et al. (2010b) compared the first two interpretations discussed above by using object-recognition tasks requiring complex within-category distinctions for several categories: *birds*, *boats*, *cars*, *chairs* and *faces*. A male patient (GG) with acquired prosopagnosia was as accurate as controls with each non-face category (see Figure 3.16). However, he was substantially less accurate than controls with faces (67% vs 94%, respectively). Thus, GG apparently has a face-specific impairment rather than a general inability to recognise complex stimuli.

Figure 3.16
Accuracy and speed of object recognition for birds, boats, cars, chairs and faces by patient GG and healthy controls.

From Busigny et al. (2012b). Reprinted with permission from Elsevier.

Busigny et al. (2010a) reviewed previous findings suggesting many patients with acquired prosopagnosia have essentially intact object recognition. However, this research was limited because the difficulty of the recognition decisions required of the patients was not controlled systematically. Busigny et al. manipulated the similarity between target items and distractors on an object-recognition task. Increasing similarity had comparable effects on PS (a patient with acquired prosopagnosia) and healthy controls. In contrast, PS performed very poorly on a face-recognition task which was very easy for healthy controls.

Why is face recognition so poor in prosopagnosics? Busigny et al. (2010b) tested the hypothesis that they have great difficulty with holistic processing. A prosopagnosic patient (GG) did not show the face inversion or composite face effects suggesting he does not perceive individual faces holistically (an ability enhancing accurate face recognition). In contrast, GG's object recognition was intact perhaps because holistic processing was not required.

Van Belle et al. (2011) also investigated the deficient holistic processing hypothesis. GG's face-recognition performance was poor when holistic processing was possible. However, it was intact when it was not possible to use holistic processing (only one part of a face was visible at a time).

Finally, we consider the expertise hypothesis. According to this hypothesis, faces differ from most other categories of objects in that we have more expertise in identifying faces. As a result, apparent differences between faces and other objects in processes and brain mechanisms may mostly reflect differences in expertise. This hypothesis is discussed further below on pp. 122–124. Barton and Corrow (2016) reported evidence consistent with this hypothesis in patients with acquired prosopagnosia who had expertise in car recognition and reading prior to their brain damage. These patients had impairments in car recognition and aspects of visual word reading suggesting they had problems with objects for which they had possessed expertise (i.e., objects of expertise). Contrary evidence was reported by Weiss et al. (2016), who studied a patient (OH) with developmental prosopagnosia. In spite of severely impaired face-recognition ability, she displayed superior recognition skills for horses (she had spent 15 years working with them). Thus, visual expertise can be acquired *independently* of the mechanisms responsible for expertise in face recognition.

In sum, the finding that many prosopagnosics have face-specific impairments is consistent with the hypothesis that face recognition involves special processes. However, more general recognition impairments have also often been reported and are apparently inconsistent with that hypothesis. There is also some support for the expertise hypothesis but again the findings are mixed.

Fusiform face area

If faces are processed differently to other objects, we would expect to find brain regions specialised for face processing. The fusiform face area (FFA) in the ventral temporal cortex has (as its name strongly implies!) been identified as such a brain region.

The fusiform face area is indisputably involved in face processing. Downing et al. (2006) found the fusiform face area responded more strongly to faces than any of 18 object categories (e.g., *tools*; *fruits*; *vegetables*). However, other brain regions, including the occipital face area (OFA) and the superior temporal sulcus (STS) are also face-selective (Grill-Spector et al., 2017; see Figure 3.17). Such findings indicate that face processing depends on one or more brain networks rather than simply on the fusiform face area.

Even though several brain areas are face-selective, the fusiform face area has been regarded as having special importance. For example, Axelrod and Yovel (2015) considered brain activity in several face-selective regions when observers were shown photos of Leonardo DiCaprio and Brad Pitt. The fusiform face area was the only region in which the pattern of brain activity differed significantly between these actors. However, Kanwisher et al. (1997) found only 80% of their participants had greater activation within the fusiform face area to faces than to other objects.

In sum, the fusiform face area plays a major role in face processing and recognition for most (but probably not all) individuals. However, face
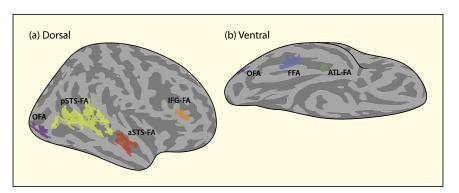
Interactive feature:
Primal Pictures'
3D atlas of the brain

Figure 3.17
Face-selective areas in the right hemisphere. OFA = occipital face area; FFA = fusiform face area; pSTS-FA and aSTS-FA = posterior and anterior superior temporal sulcus face areas; IFG-FA = inferior frontal gyrus face area; ATL-FA = anterior temporal lobe face area.
From Duchaine and Yovel (2015).

processing depends on a brain network, including several areas in addition to the fusiform face area (see Figure 3.17). Note also that the fusiform face area is activated when we process numerous types of non-face objects. Finally, face-processing deficits in prosopagnosics are *not* limited to the fusiform face area. For example, developmental prosopagnosics had less selectivity for faces than healthy controls in 12 different face areas (including the fusiform face area) (Jiahui et al., 2018).

Expertise hypothesis

According to advocates of the expertise hypothesis (e.g., Wang et al., 2016; discussed on p. 119), major differences between face and object processing should not be taken at face value (sorry!). According to this hypothesis, the brain and processing mechanisms allegedly specific to faces are also involved in processing and recognising all object categories for which we possess *expertise*. Thus, we should perhaps relabel the fusiform face area as the "fusiform expertise area".

Why is expertise so important in determining face and object processing? One reason is that expertise leads to greater holistic or integrated processing. For example, chess experts can very rapidly use holistic processing based on their relevant stored knowledge to understand complex chess positions (see Chapter 12).

Three main predictions follow from the expertise hypothesis:

(1) Holistic or configural processing is *not* unique to faces but should be found for any objects of expertise.
(2) The fusiform face area should be highly activated when observers recognise the members of any category for which they possess expertise.
(3) If the processing of faces and of objects of expertise involves similar processes, then processing objects of expertise should interfere with face processing.

Findings

The first prediction is plausible. Wallis (2013) tested a model of object recognition to assess the effects of prolonged exposure to any given stimulus category. The model predicted that many phenomena associated with face processing (e.g., holistic processing; inversion effect) would be found for *any* stimulus category for which observers had expertise. Repeated simultaneous presentation of the same features (e.g., nose; mouth; eyes) gradually increases holistic processing. Wallis concluded a single model can explain object *and* face recognition.

There is some support for the first prediction in research on detection of abnormalities in medical images (see Chapter 12). Kundel et al. (2007) found experts generally fixated on an abnormality in under 1 second suggesting they used very fast, holistic processes. However, as we saw earlier, experts with non-face objects often have a small inversion effect (assumed to reflect holistic processing). McKone et al. (2007) found such experts

rarely show the composite effect (also assumed to reflect holistic processing; discussed on p. 117).

We turn now to the second prediction. In a review, McKone et al. (2007) found a modest tendency for the fusiform face area to be more activated by objects of expertise than other objects. However, larger activation effects for objects of expertise were found *outside* the fusiform face area than *inside* it. Support for the second prediction was reported by McGugin et al. (2014): activation to car stimuli within the fusiform face area was greater in participants having greater car expertise.

McGugin et al. (2018) argued that we can test the second prediction by comparing individuals varying in face-recognition ability (or expertise). As predicted, those with high face-recognition ability exhibited more face-selective activation within the fusiform face area than those having low ability.

Bilalić (2016) found chess experts had more activation in the fusiform face area than non-experts when viewing chess positions but *not* single chess pieces. He concluded, "The more complex the stimuli, the more likely it is that the brain will require the help of the FFA in grasping its essence" (p. 1356).

It is important not to oversimplify the issues here. Even if face processing and processing of other objects of expertise both involve the fusiform face area, the two forms of processing may use different neurons in different combinations (Grill-Spector et al., 2017).

We turn now to the third prediction. McKeeff et al. (2010) found that car experts were slower than novices when searching for face targets among cars but not among watches. Car and face expertise may have *interfered* with each other because they depend on similar processes. Alternatively, car experts may have been more likely than car novices to attend to distracting cars because they find such stimuli more interesting.

McGugin et al. (2015) also tested the third prediction. Overall, car experts had greater activation than car novices in face-selective areas (e.g., fusiform face area) when processing cars. Of key importance, that difference was greatly reduced when faces were also presented. Thus, interference was created when participants processed objects belonging to two different categories of expertise (i.e., cars and faces).

Evaluation

There is some support for the expertise hypothesis with respect to all three predictions. However, the extent of that support remains controversial. One reason is that it is hard to assess expertise level accurately or to control it. It is certainly possible that many (but not all) processing differences between faces and other objects are due to greater expertise with faces. This would imply that faces are less special than often assumed.

According to the expertise hypothesis, we are face experts. This may be true of familiar faces, but it is certainly *not* true of unfamiliar faces (Young & Burton, 2018). Evidence of the problems we experience in recognising unfamiliar faces is contained in the Box on passport control.

IN REAL LIFE: PASSPORT CONTROL 🔍

Look at the 40 faces displayed below (see Figure 3.18). How many different individuals are shown? Provide your answer before reading on.

Figure 3.18
An array of 40 face photographs to be sorted into piles for each of the individuals shown in the photographs.
From Jenkins et al. (2011). Reproduced with permission from the Royal Society.

In a study by Jenkins et al. (2011) using a similar stimulus array, participants on average decided 7.5 different individuals were shown. However, the actual number for the array used by Jenkins et al. and the one shown in Figure 3.18 is only two! The two individuals (A and B) are arranged as shown below:

A B A A A B A B A B

A A A A A B B B A B

B B B A A A B B A A

B A B A A B B B B B

Perhaps we are poor at matching unfamiliar faces because we rarely perform this task in everyday life. White et al. (2014) addressed this issue in a study on passport officers averaging 8 years of service. These passport officers indicated on each trial whether a photograph was that of a physically present person. Overall, 6% of valid photos were rejected and 14% of fraudulent photos were wrongly accepted. Thus, individuals with specialist training and experience are *not* exempt from problems in matching unfamiliar faces. The main problem is that there is considerable *variability* in how an individual looks in different photos (discussed further on p. 127).

In another experiment, White et al. (2014) compared the performance of passport officers and students on a matching task with unfamiliar faces. The two groups were comparable with 71% correct performance on match trials and 89% on non-match trials. Thus, training and experience were irrelevant.

In White et al.'s (2014) research, 50% of the photos were invalid (non-matching). This is (hopefully!) a massively higher percentage of invalid photos than typically found at passport control. Papesh and Goldinger (2014) compared performance when actual mismatches occurred on 50% or 10% of trials. In the 50% condition, mismatches were missed on 24% of trials, whereas they were missed on 49% of trials in the 10% condition. Participants had a low expectation of mismatches in the 10% condition and so were very cautious about deciding two photos were of different individuals (i.e., they had a very cautious response criterion indicating response bias).

Papesh et al. (2018) replicated the above findings. They attempted to improve performance in the condition where mismatches occurred on only 10% of trials by introducing blocks where mismatches occurred on 90% of trials. However, this manipulation had little effect because participants were reluctant to abandon their very cautious response criterion.

How can we provide better security at passport control? Increased practice at matching unfamiliar faces is *not* the answer – White et al. (2014) found performance was unrelated to the number of years passport officers had served. A promising approach is to find individuals having an exceptional ability to recognise faces (**super-recognisers**). Robertson et al. (2016) asked participants to decide whether face pairs depicted the same person. Mean accuracy was 96% for previously identified police super-recognisers compared to only 81% for police trainees.

Why do some individuals have very superior face-recognition ability? Wilmer et al. (2010) found the face-recognition performance of monozygotic (identical) twins was much closer than that of dizygotic (fraternal) twins, indicating face-recognition ability is strongly influenced by genetic factors. Face-recognition ability correlated very modestly with other forms of recognition (e.g., abstract art images), suggesting it is very specific. In similar fashion, Turano et al. (2016) found good and poor face recognisers did not differ with respect to car-recognition ability.

Theoretical approaches

Bruce and Young's (1986) model has been the most influential theoretical approach to face processing and recognition and so we start with it. It is a serial stage model consisting of eight components (see Figure 3.19):

Interactive exercise:
Face recognition

(1) *Structural encoding*: this produces various descriptions or representations of faces.
(2) *Expression analysis*: people's emotional states are inferred from their facial expression.
(3) *Facial speech analysis*: speech perception is assisted by lip reading (see Chapter 9).
(4) *Direct visual processing*: specific facial information is processed selectively.
(5) *Face recognition units*: these contain structural information about known faces; this structural information emphasises the less changeable aspects of the face and is fairly abstract.
(6) *Person identity nodes*: these provide information about individuals (e.g., occupation; interests).
(7) *Name generation*: a person's name is stored separately.

> **KEY TERM**
>
> **Super-recognisers**
> Individuals with an outstanding ability to recognise faces.
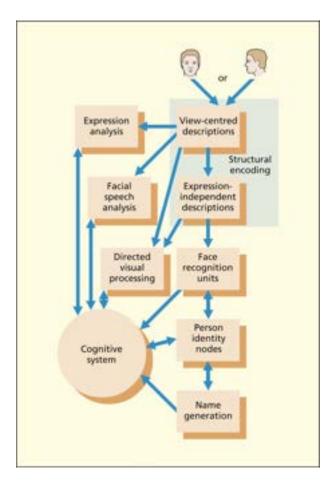
Figure 3.19
The model of face recognition put forward by Bruce and Young (1986).

Adapted from Bruce and Young (1986). Reprinted with permission of Elsevier.

(8) *Cognitive system*: this contains additional information (e.g., most actors have attractive faces); it influences which components receive attention.

What predictions follow? First, there should be major differences in the processing of familiar and unfamiliar faces because various components (face recognition units; person identity nodes; name generation) are involved only when processing familiar faces. Thus, it is much easier to recognise familiar faces, especially when faces are seen from an unusual angle.

Second, *separate* processing routes are involved in working out facial identity (who is it?) and facial expression (what is he/she feeling?). The former processing route (including the occipital face area and the fusiform face area) focuses on relatively unchanging aspects of faces, whereas the latter (involving the superior temporal sulcus) deals with more changeable aspects. This separation between processes responsible for recognising identity and expression makes sense – if there were no separation, we would have great problems recognising familiar faces with unusual expressions (Young, 2018).

Third, when we see a familiar face, familiarity information from the face recognition unit should be accessed first. This is followed by information about that person (e.g., occupation) from the person identity node and then that person's name from the name generation component. As a result, we can find a face familiar while unable to recall anything else about that person, or we can recall personal information about a person while being unable to recall their name. However, a face should *never* lead to recall of the person's name in the absence of other information.

Fourth, the model assumes face processing involves several stages. This implies the nature of face-processing impairments in brain-damaged patients depends on *which* stages of processing are impaired. Davies-Thompson et al. (2014) developed the model to account for three forms of face impairment (see Figure 3.20).

Findings

According to the model, it is easier to recognise familiar faces than unfamiliar ones for various reasons. Of special importance, we possess much more structural information about familiar faces. This structural information
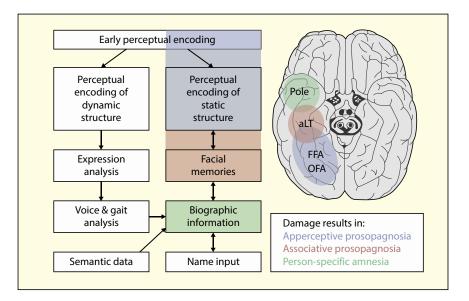
Figure 3.20
Damage to regions of the inferior occipito-temporal cortex (including fusiform face area (FFA) and occipital face area (OFA)) is associated with apperceptive prosopagnosia (blue); damage to anterior inferior temporal cortex (aLT) is associated with associative prosopagnosia (red); and damage to the anterior temporal pole is associated with person-specific amnesia (green). Davies-Thompson et al. (2014) discuss evidence consistent with their model.

From Davies-Thompson et al. (2014).

(associated with face recognition units) relates to relatively unchanging aspects of faces and gradually accumulates with increasing familiarity with any given face.

However, the differences in ease of recognition between familiar and unfamiliar faces are greater than envisaged by Bruce and Young (1986). Jenkins et al. (2011) found 40 face photographs showing only two different unfamiliar individuals were thought to show almost four times that number (discussed on p. 124). The two individuals were actually two Dutch celebrities almost unknown in Britain. When Jenkins et al. (2011) repeated their experiment with Dutch participants, nearly all performed the task perfectly because the faces were so familiar.

Why is unfamiliar face recognition so difficult? There is considerable within-person variability in facial images, which is why different photographs of the same unfamiliar individual often look as if they come from different individuals (Young & Burton, 2017, 2018). Jenkins and Burton (2011) argued we could improve identification of unfamiliar faces by *averaging* across several photographs of the same individual and so greatly reducing image variability. Their findings supported this prediction.

Burton et al. (2016) shed additional light on the complexities of recognising unfamiliar faces. In essence, how one person's face varies across images differs from how someone else's face varies. Thus, the characteristics that vary or remain constant across images differ from one individual to another.

The second prediction is that different routes are involved in the processing of facial identity and facial expression. There is some support for this prediction. Fox et al. (2011) found patients with damage to the face-recognition network had impaired identity perception but not expression perception. In contrast, a patient with damage to the superior temporal sulcus had impaired expression perception but reasonably intact identity perception. Sliwinska and Pitcher (2018) confirmed the role played by the superior temporal sulcus. Transcranial magnetic stimulation (TMS; see Glossary) applied to this area impaired recognition of facial expression.

However, the two routes are not entirely independent. Judgements of facial expression are strongly influenced by irrelevant identity information (Schweinberger & Soukup, 1998). Redfern and Benton (2017) asked participants to sort cards of faces into piles, one for each perceived identity. One pack contained expressive faces and the other neutral faces. With expressive faces, faces belonging to different individuals were more likely to be placed in the same pile. Thus, expressive facial information can influence (and impair) identity perception.

Fitousi and Wenger (2013) asked participants to respond positively to a face that had a given identity and emotion (e.g., a happy face belonging to Kiera Knightley). Facial identity and facial expression were *not* processed independently although they should have been according to the model.

Another issue is that the facial expression route is more complex than assumed theoretically. For example, damage to the amygdala produces greater deficits in recognising fear and anger than other emotions (Calder & Young, 2005). Young and Bruce (2011) admitted they had not expected deficits in emotion recognition in faces to vary across emotions.

The third prediction is that we always retrieve personal information (e.g., occupation) about a person *before* recalling their name. Young et al. (1985) asked people to record problems they experienced in face recognition. There were 1,008 such incidents but people never reported putting a name to a face while knowing nothing else about that person. In contrast, there were 190 occasions on which someone remembered a reasonable amount of information about a person but not their name (also as predicted by the model).

Several other findings support the third prediction (Hanley, 2011). However, the notion that names are *always* recalled after personal information is too rigid. Calderwood and Burton (2006) asked fans of the television series *Friends* to recall the name or occupation of the main characters when shown their faces. Names were recalled faster than occupations (against the model's prediction).

Fourth, we relate face-processing impairments to Bruce and Young's (1986) serial stage model. We consider three such impairments (discussed by Davies-Thompson et al., 2014; see Figure 3.20) with reference to Figure 3.19:

(1) *Patients with impaired early stages of face processing*: such patients (categorised as having *apperceptive prosopagnosia*) have "an inability to form a sufficiently accurate representation of the face's structure from visual data" (Davies-Thompson et al., 2014, p. 161). As a result, faces are often not recognised as familiar.

(2) *Patients with impaired ability to access facial memories in face rec-ognition units although early processing of facial structure is relatively intact*: such patients have *associative prosopagnosia*: they have greater problems with memory than perception.

(3) *Patients with impaired access to biographical information stored in person identity nodes*: such patients have *person-specific amnesia* and differ from those with associative prosopagnosia because they often cannot recognise other people by *any* cues (including spoken names or voices).

So far we have applied Bruce and Young's (1986) model to acquired prosop-agnosia. However, we can also apply the model to developmental prosop-agnosia (in which face-recognition mechanisms fail to develop normally). Parketny et al. (2015) presented previously unfamiliar faces to develop-mental prosopagnosics and recorded event-related potentials (ERPs) while they performed an easy face-recognition task. They focused on three ERP components:

(1) N170: this early component (about 170 ms) reflects processes involved in perceptual structural face processing.

(2) N250: this component (about 250 ms) reflects a match between a pre-sented face and a stored face representation.

(3) P600: this component (about 600 ms) reflects attentional processes associated with face recognition.

What did Parketny et al. (2015) find? Recognition times were 150 ms slower in the developmental prosopagnosics than healthy controls. N170 was broadly similar in both groups with respect to timing and magnitude. N250 was 40 ms slower in the prosopagnosics than controls but of comparable magnitude. Finally, P600 was significantly smaller in the prosopagnosics than controls and was delayed by 80 ms. In sum, developmental prosop-agnosics show relatively intact early face processing but are slower and less efficient later in processing. ERPs provide an effective way of identifying those aspects of face processing adversely affected in prosopagnosia.

Evaluation

Bruce and Young (1986) provided a comprehensive framework emphasis-ing the wide range of information that can be extracted from faces. It was remarkably innovative in identifying the major processes and structures involved in face processing and recognition and incorporating them within a plausible serial stage approach. Finally, the model enhanced our understand-ing of why familiar faces are much easier to recognise than unfamiliar ones.

What are the model's limitations? First, the complexities involved in recognising unfamiliar faces (e.g., coping with the great variability in a given individual's facial images) were not fully acknowledged. As Young and Burton (2017, p. 213) pointed out, it was several years after 1986 before researchers appreciated that "humans' relatively poor performance at unfamiliar-face recognition is as much a problem of perception as of memory".

Case study:
Model of face processing

Second, the model's account of the processing of facial expression is oversimplified. For example, the processing of facial expression is less independent of the processing of facial identity than assumed theoretically. According to the model, damage to the expression analysis component should produce impaired ability to recognise all facial expressions. In fact, many brain-damaged patients have much greater impairment in facial recognition of some emotions than others (Young, 2018).

Third, the model was somewhat vague about the precise information stored in the face recognition units and the person identity nodes. Fourth, it was wrong to exclude gaze perception from the model because it provides useful information about what an observer is attending to (Young & Bruce, 2011). Fifth, Bruce and Young (1986) focused on general factors influencing face recognition. However, as discussed earlier, there are substantial individual differences in face-recognition ability with a few individuals (super-recognisers) having outstanding ability. These individual differences depend mostly on genetic factors (Wilmer, 2017) not considered within the model.

VISUAL IMAGERY

Close your eyes and imagine the face of someone you know very well. What did you experience? Many people claim forming visual images is like "seeing with the mind's eye", suggesting there are important similarities between imagery and perception. Mental imagery is typically regarded as involving conscious experience. However, we could also regard imagery as a form of mental representation (an internal cognitive symbol representing aspects of external reality) (e.g., Pylyshyn, 2002). We would not necessarily be consciously aware of images as mental representations.

Galton (1883) supported the above viewpoint. He found many individuals reported no conscious imagery when imagining a definite object (e.g., their breakfast table). Zeman et al. (2015) studied several individuals lacking visual imagery and coined the term aphantasia to refer to this condition.

If visual imagery and perception are similar, why do we very rarely confuse them? One reason is that we are generally aware of *deliberately* constructing images (unlike with visual perception). Another reason is that images contain much less detail. For example, people rate their visual images of faces as similar to photographs lacking sharp edges and borders (Harvey, 1986).

However, many people sometimes confuse visual imagery and perception. Consider hallucinations in which perception-like experiences occur in the absence of the appropriate environmental stimulus. Visual hallucinations occur in approximately 27% of schizophrenic patients but also in 7% of the general population. Waters et al. (2014) discussed research showing visual hallucinations in schizophrenics are often associated with activity in the primary visual cortex, suggesting hallucinations involve many processes associated with visual perception. One reason schizophrenics are susceptible to visual hallucinations is because of distortions in top-down processing (e.g., forming strong expectations of what they will see).

In **Anton's syndrome** ("blindness denial"), blind people are unaware that they are blind and sometimes confuse imagery for actual perception. Goldenberg et al. (1995) described a patient whose primary visual cortex had been nearly wholly destroyed. Nevertheless, she generated such vivid visual images that she mistook them for genuine visual perception. The brain damage in patients with Anton's syndrome typically includes large parts of the visual cortex (Gandhi et al., 2016).

There is also **Charles Bonnet syndrome**, defined as "consistent or periodic complex visual hallucinations that occur in visually impaired individuals" (Yacoub & Ferrucci, 2011, p. 421). However, patients are generally aware the hallucinations are not real and so they are actually pseudo-hallucinations. When patients hallucinate, they have increased activity in brain areas specialised for visual processing (e.g., hallucinations in colour are associated with activity in colour-processing areas) (ffytche et al., 1998).

Painter et al. (2018) identified a major reason for this elevated activity. Stimuli presented to intact regions of the retina cause extreme excitability (hyperexcitability) within early visual cortex. Visually impaired individuals with hallucinations show greater hyperexcitability than those without.

Why is visual imagery useful?

What functions are served by visual imagery? According to Moulton and Kosslyn (2009, p. 1274), visual imagery "allows us to answer 'what if' questions by making explicit and accessible the likely consequences of being in a specific situation or performing a specific action". For example, professional golfers use mental imagery to *predict* what would happen if they hit a certain shot.

Pearson and Kosslyn (2015) pointed out that many visual images contain rich information that is accessible when required. For example, what is the shape of a cat's ears? You may be able to answer the question by constructing a visual image.

More generally, visual imagery supports numerous cognitive functions. These include creative insight, attentional search, guiding deliberate action, short-term memory storage and long-term memory retrieval (Mitchell & Cusack, 2016).

Imagery theories

Kosslyn (e.g., 1994; Pearson & Kosslyn, 2015) proposed an influential theory based on the assumption that visual imagery resembles visual perception. It was originally called perceptual anticipation theory because image generation involves processes used to anticipate perceiving visual stimuli.

According to the theory, visual images are depictive representations. What is a **depictive representation**? In such a depiction, "each part of the representation corresponds to a part of the represented object such that the distances among the parts in the representation correspond to the actual distances among the parts" (Pearson & Kosslyn, 2015, p. 10089). Thus, for example, a visual image of a desk with a computer on top and a cat

KEY TERMS

Anton's syndrome
A condition found in some blind people in which they misinterpret their visual imagery as visual perception.

Charles Bonnet syndrome
A condition in which individuals with eye disease form vivid and detailed visual hallucinations sometimes mistaken for visual perception.

Depictive representation
A representation (e.g., visual image) resembling a picture in that objects within it are organised spatially.

Research activity:
Mental imagery

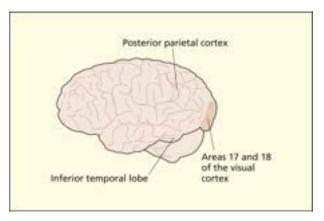**Interactive exercise:**
Kosslyn – mental imagery

Figure 3.21
The approximate locations of the visual buffer in BA17 and BA18, of long-term memories of shapes in the inferior temporal lobe, and of spatial representations in the posterior parietal cortex, according to Kosslyn and Thompson's (2003) anticipation theory.

sleeping underneath would have the computer at the top and the cat at the bottom.

Where are depictive representations formed? Kosslyn argued they are created in early visual cortex (BA17 and BA18; see Figure 3.21) within a visual buffer. The **visual buffer** is a short-term store for visual information only and is of major importance in visual perception and imagery. There is also an "attention window" selecting some visual information in the visual buffer and passing it on to other brain areas for further processing. This attention window is *flexible* – it can be adjusted to include more, or less, visual information.

Processing in the visual buffer depends primarily on external stimulation during perception. However, such processing involves non-pictorial information stored in long-term memory during imagery. Shape information is stored in the inferior temporal lobe whereas spatial representations are stored in posterior parietal cortex (see Figure 3.21). In sum, visual perception mostly involves bottom-up processing whereas visual imagery depends on top-down processing.

Pylyshyn (e.g., 2002) argued visual imagery differs substantially from visual perception. According to his propositional theory, performance on mental imagery tasks does *not* involve depictive or pictorial representations. Instead, it involves tacit knowledge (knowledge inaccessible to conscious awareness). Tacit knowledge is "Knowledge of *what things would look like* to subjects in situations like the ones in which they are to imagine themselves" (Pylyshyn, 2002, p. 161). Thus, performance on an imagery task relies on relevant stored knowledge rather than visual images. Within this theoretical framework, it is improbable that early visual cortex would be involved on an imagery task.

Imagery resembles perception

KEY TERMS 🔑

Visual buffer
Within Kosslyn's theory, a short-term visual memory store involved in visual imagery and perception.

Binocular rivalry
When two different visual stimuli are presented one to each eye, only one stimulus is seen; the seen stimulus alternates over time.
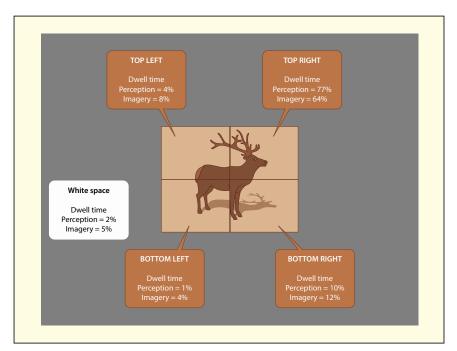
If visual perception and imagery involve similar processes, they should influence each other. There should be *facilitation* if the contents of perception and imagery are the same but *interference* if they differ. Pearson et al. (2008) reported a facilitation effect with **binocular rivalry** – when a different stimulus is presented to each eye, only one is consciously perceived at any given moment. The act of imagining a specific pattern strongly influenced which stimulus was subsequently perceived and this facilitation depended on the similarity between the imagined and presented stimuli. The findings were remarkably similar when the initial stimulus was perceived rather than imagined.

Baddeley and Andrade (2000) reported an interference effect. Participants rated the vividness of visual and auditory images while performing a second task involving visual/spatial processes. This task reduced

the vividness of *visual* imagery more than that of *auditory* imagery because similar processes were involved on the visual/spatial and visual imagery tasks.

Laeng et al. (2014) asked participants to view pictures of animals and to follow each one by forming a visual image of that animal. There was a striking similarity in eye fixations devoted to the various areas of each picture in both conditions (see Fig. 3.22). Participants having the greatest similarity in dwell time between perception and imagery showed the best memory for the size of each animal.

According to Kosslyn's theoretical position, much processing associated with visual imagery occurs in early visual cortex (BA17 and BA18) plus several other areas. In a review, Kosslyn and Thompson (2003) found 50% of studies using visual-imagery tasks reported activation in early visual cortex. Significant findings were most likely when the task involved inspecting the fine details of images or focusing on an object's shape. In a meta-analysis (see Glossary), Winlove et al. (2018) found the early visual cortex (V1) was typically activated during visual imagery. Consistent with Kosslyn's theory, activation in the early visual cortex is greater among individuals reporting vivid visual imagery.

The neuroimaging evidence discussed above is limited – it is correlational and so the activation associated with visual imagery may not be *directly* relevant to the images that are formed. Naselaris et al. (2015) reported more convincing evidence. Participants formed images of five artworks. It was possible to some extent to identify the imagined artworks from hundreds of other artworks through careful analysis of activity in the early visual cortex. Some of this activity corresponded to the processing of low-level visual features (e.g., space; orientation).

Further neuroimaging support for the notion that imagery closely resembles perception was reported by Dijkstra et al. (2017a). They found "the overlap in neural representations between imagery and perception . . . extends beyond the visual cortex to include also parietal and premotor/frontal areas" (p. 1372). Of most importance, the greater the neural overlap between imagery and perception throughout the entire visual system, the more vivid was the imagery experience.

Imagery does *not* resemble perception

Look at Figure 3.23. Start with the object on the left and form a clear image of it. Then close your eyes, mentally rotate the image by 90º clockwise and decide what you see. Then repeat the exercise with the other objects. Finally, rotate the book through 90º. You probably found it very easy to identify the objects when perceiving them but impossible when only imagining rotating them. Slezak (1991, 1995) used stimuli closely resembling those in Figure 3.23 and found no observers reported seeing the objects. Thus, the information within images is much less detailed and flexible than visual information.

Lee et al. (2012) identified important differences between imagery and perception. Observers viewed or imagined common objects (e.g., *car*; *umbrella*) while activity in the early visual cortex and areas associated with later visual processing (object-selective regions) was assessed. Attempts were made by the researchers to identify the objects being imagined or perceived on the basis of activation in those areas.

What did Lee et al. (2012) find? First, activation in all brain areas was considerably greater when participants perceived rather than imagined objects. Second, objects being perceived or imagined were identified with above-chance accuracy based on patterns of brain activation except for imagined objects in the primary visual cortex (V1; see Figure 3.24).
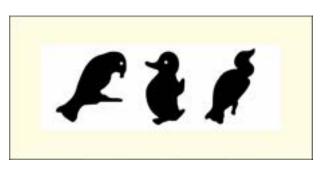
Third, the success rate in identifying *perceived* objects was greater based on brain activation in areas associated with *early* visual processing than those associated with later processing. However, the opposite was the case with *imagined* objects (see Figure 3.24). Thus, object processing in the early visual cortex is very limited during imagery but is extremely important during perception. Imagery for objects depends mostly on top-down processes based on object knowledge rather than processing in the early visual cortex.

Most cognitive neuroscience research has focused on the brain areas activated during visual perception and imagery. It is also important to focus on *connectivity* between brain areas. Dijkstra et al. (2017b) considered connectivity among four brain areas of central importance in perception and imagery: early
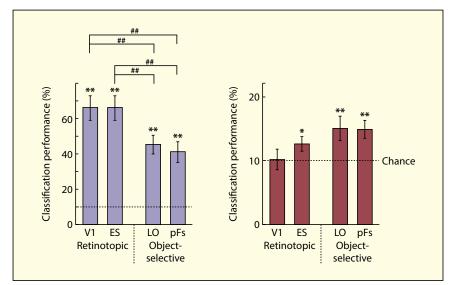
Figure 3.23

Slezak (1991, 1995) asked participants to memorise one of the above images. They then imagined rotating the image 90 degrees clockwise and reported what they saw. None of them reported seeing the figures that can be seen clearly if you rotate the page by 90 degrees clockwise.

Left image from Slezak (1995), centre image from Slezak (1991), right image reprinted from Pylyshyn (2002), with permission from Elsevier and the author.

Figure 3.24

The extent to which perceived (left side of figure) or imagined (right side of figure) objects could be classified accurately on the basis of brain activity in the early visual cortex and object-selective cortex. ES =extrastriate retinotopic cortex; LO = lateral occipital cortex; pFs = posterior fusiform sulcus.

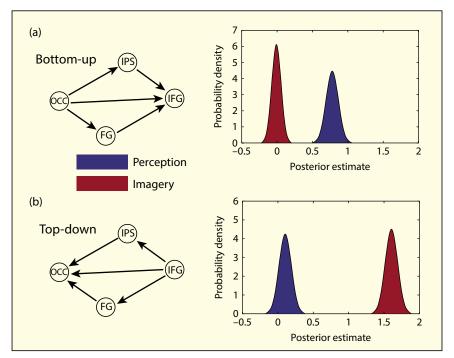From S.H. Lee et al. (2012). Reproduced with permission from Elsevier.

Figure 3.25

Connectivity during perception and imagery involving (a) bottom-up processing; and (b) top-down processing. Posterior estimates indicate connectivity strength (the further from 0 the stronger). The meanings of OCC, FG, IPS and IFG are given in the text.

From Dijkstra et al. (2017b).

visual cortex (OCC), fusiform gyrus (FG; late visual cortex), IPS (intraparietal sulcus) and IFG (inferior frontal gyrus). The first two are mostly associated with bottom-up processing whereas the second two are mostly associated with top-down processing.

Dijkstra et al.'s (2017b) key findings are shown in Figure 3.25. First, perception was associated with reasonably strong bottom-up brain connectivity and weak top-down brain connectivity. Second, imagery was associated with non-significant bottom-up connectivity but very strong top-down connectivity. Thus, top-down connectivity from frontal to early visual areas

is a common mechanism during perception and imagery. However, there is much stronger top-down connectivity during imagery to compensate for the absence of bottom-up connectivity. Individuals having the greatest top-down connectivity during imagery reported the most vivid images.

Dijkstra et al. (2018) studied the time course for the development of visual representations in perception and in imagery using magneto-encephalography (MEG; see Glossary). With perception, they confirmed that visual representations develop through a series of processing stages (see Chapter 2). With imagery, in contrast, the entire visual representation appeared to be activated simultaneously, presumably because all the relevant information was retrieved together from memory.

Brain damage

If visual perception and visual imagery involve the same mechanisms, we might expect brain damage to have comparable effects on perception and imagery. That is often the case. However, there are numerous exceptions (Bartolomeo, 2002, 2008). Moro et al. (2008) studied two brain-damaged patients with intact visual perception but impaired visual imagery. They were both very poor at drawing objects from memory but could copy the same objects when shown a drawing.

These patients (and others with impaired visual imagery but intact visual perception) have damage to the left temporal lobe. Visual images are probably generated from information about concepts (including objects) stored in the temporal lobes (Patterson et al., 2007). However, this generation process is less important for visual perception.

Bridge et al. (2012) studied a young man, SBR, who had virtually no primary visual cortex and nearly total blindness. However, he had vivid visual imagery and his pattern of cortical activation when engaged in visual imagery resembled that of healthy controls. Similar findings were reported with a 70-year-old woman, SH, who became blind at the age of 27. She had intact visual imagery predominantly involving areas outside the early visual cortex. Of relevance, she had greater connectivity between some visual networks in the brain than most individuals.

How can we interpret the above findings? Visual perception mostly involves bottom-up processes triggered by the stimulus whereas visual imagery primarily involves top-down processes based on object knowledge. Thus, it is unsurprising brain areas involved in early visual processing are more important for perception than imagery whereas brain areas associated with storage of information about visual objects are more important for imagery.

Evaluation

Much progress has been made in understanding the relationship between visual imagery and visual perception. Similar processes are involved in imagery and perception and they are both associated with somewhat similar patterns of brain activity. In addition, the predicted facilitatory and interfering effects between imagery and perception tasks have been reported. These findings are more consistent with Kosslyn's theory than Pylyshyn's.

On the negative side, visual perception and visual imagery are less similar than assumed by Kosslyn. For example, there is the neuroimaging evidence reported by Lee et al. (2012) and the frequent dissociations between perception and imagery found in brain-damaged patients. Of most importance, visual perception involves strong bottom-up connectivity and weak top-down connectivity, whereas visual imagery involves very strong top-down connectivity but negligible bottom-up connectivity (Dijkstra et al., 2017b).

CHAPTER SUMMARY

- **Pattern recognition.** Pattern recognition involves processing of specific features and global processing. Feature processing generally (but not always) precedes global processing. Several types of cells (e.g., simple cells; complex cells; end-stopped cells) are involved in feature processing. There are complexities in pattern recognition due to interactions among cells and the influence of top-down processes. Evidence from computer programs to solve CAPTCHAs suggests humans are very good at processing edge corners. Fingerprint identification is sometimes very accurate; however, even experts show confirmation bias (distorted performance caused by contextual information). Fingerprint experts are much better than novices at discriminating between matches and non-matches and also adopt a more conservative response bias.

- **Perceptual organisation.** The gestaltists proposed several principles of perceptual grouping and emphasised the importance of figure-ground segmentation. They argued that perceptual grouping and figure-ground segregation depend on innate factors. They also argued we perceive the simplest possible organisation of the visual field. The gestaltists provided descriptions rather than explanations. Their approach underestimated the complex interactions of factors underlying perceptual organisation. The gestaltists de-emphasised the role of experience and learning in perceptual organisation. However, recent theories based on Bayesian inference (e.g., the Bayesian hierarchical grouping model) have emphasised learning processes and fully acknowledge the importance of learning.

- **Approaches to object recognition.** Visual processing typically involves a coarse-to-fine processing sequence: low spatial frequencies in visual input (associated with coarse processing) are conveyed to higher visual areas faster than high spatial frequencies (associated with fine processing). Biederman assumed in his recognition-by-components theory that objects consist of geons (basic shapes). An object's geons are determined by edge-extraction processes and the resultant geon-based description is viewpoint-invariant. Biederman's theory de-emphasises the role of top-down processes.

Object recognition is sometimes viewpoint-invariant (as predicted by Biederman) with easy categorical discriminations, but it is more typically viewer-centred when identification is required. Object representations often contain viewpoint-dependent *and* viewpoint-invariant information.

- **Object recognition: top-down processes.** Top-down processes are more important in object recognition when observers view degraded or briefly presented stimuli. Top-down processes sometimes influence attention, memory or response bias rather than perception itself. However, there are also direct effects of top-down processes on object recognition. According to the interactive-iterative framework (Baruch et al., 2018), top-down and bottom-up processes interact with top-down processes (e.g., attention) influencing subsequent bottom-up processing.

- **Face recognition.** Face recognition involves more holistic processing than object recognition. Deficient holistic processing partly explains why prosopagnosic patients have much greater problems with face recognition than object recognition. Face processing involves a brain network including the fusiform face and occipital face areas. However, much of this network is also used in processing other objects (especially when recognising objects for which we have expertise).

  Bruce and Young's model assumes several serial processing stages. Research on prosopagnosics supports this assumption because the precise nature of their face-recognition impairments depends on which stage(s) are most affected. The model also assumes there are major differences in the processing of familiar and unfamiliar faces. This assumption has received substantial support. However, Bruce and Young did not fully appreciate that unfamiliar faces are hard to recognise because of the great variability of any given individual's facial images. The model assumes there are two independent processing routes (for facial expression and facial identity), but they are not entirely independent. The model ignores the role played by genetic factors in accounting for individual differences in face-recognition ability.

- **Visual imagery.** Visual imagery allows us to predict the visual consequences of performing certain actions. According to Kosslyn's perceptual anticipation theory, visual imagery closely resembles visual perception. In contrast, Pylyshyn, in his propositional theory, argued visual imagery involves making use of tacit knowledge and does not resemble visual perception. Visual imagery and perception influence each other as predicted by Kosslyn's theory. Neuroimaging studies and studies on brain-damaged patients indicate similar areas are involved in imagery

and perception. However, areas involved in top-down processing (e.g., left temporal lobe) are more important in imagery than perception, and areas involved in bottom-up processing (e.g., early visual cortex) are more important in perception. More generally, bottom-up brain connectivity is far more important in perception than imagery, whereas top-down brain connectivity is far more important in imagery than perception.

FURTHER READING

Baruch, O., Kimchi, R. & Goldsmith, M. (2018). Attention to distinguishing features in object recognition: An interactive-iterative framework. *Cognition*, *170*, 228–244. Orit Baruch and colleagues provide a theoretical framework for understanding how bottom-up and top-down processes interact in object recognition.

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M.A.J. & Friston, K. (2017b). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports*, *7* (Article 5677). In this article, Nadine Dijkstra and her colleagues clarify the roles of top-down and bottom-up processes in visual perception and imagery.

Firestone, C. & Scholl, B.J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, 1–77. The authors argue that top-down processes do not directly influence visual perception. Read the open peer commentary following the article, however, and you will see most experts disagree.

Gauthier, I. & Tarr, M.J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science*, *2*, 377–396. Isabel Gauthier and Michael Tarr provide a comprehensive overview of theory and research on object recognition.

Grill-Spector, K., Weiner, K.S., Kay, K. & Gomez, J. (2017). The functional neuroanatomy of human face perception. *Annual Review of Vision Science*, *3*, 167–196. This article by Kalanit Grill-Sector and colleagues contains a comprehensive account of brain mechanisms underlying face perception.

Wagemans, J. (2018). Perceptual organisation. In J.T. Serences (ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. 2: Sensation, Perception, and Attention* (4th edn; pp. 803–822). New York: Wiley. Johan Wagemans reviews various theoretical and empirical approaches to understanding perceptual organisation.

Young, A.W. (2018). Faces, people and the brain: The 45th Sir Frederic Bartlett lecture. *Quarterly Journal of Experimental Psychology*, *71*, 569–594. Andy Young provides a very interesting account of theory and research on face perception.

Chapter

4

Motion perception and action

INTRODUCTION

Most research on perception discussed in previous chapters involved presenting a visual stimulus and assessing aspects of its meaning. What was missing (but is an overarching theme of this chapter) is the *time* dimension. In the real world, we move around and/or people or objects in the environment move. The resulting changes in the visual information available to us are very useful in ensuring we perceive the environment accurately and also respond appropriately. This emphasis on *change* and *movement* necessarily leads to a consideration of the relationship between perception and action. In sum, our focus in this chapter is on how we process (and respond to) a constantly changing environment.

The first theme addressed in this chapter is the perception of movement. This includes our ability to move successfully within the visual environment and predict accurately when moving objects will reach us.

The second theme is concerned with more complex issues – how do we act appropriately on the environment and the objects within it? Of relevance are theories (e.g., the perception-action theory; the dual-process approach) distinguishing between processes and systems involved in vision-for-perception and those involved in vision-for-action (see Chapter 2). Here we consider theories providing more detailed accounts of vision-for-action and/or the workings of the dorsal pathways allegedly underlying vision-for-action.

The third theme focuses on the processes involved in *making sense* of moving objects (especially other people). It thus differs from the first theme in which moving stimuli are considered mostly in terms of predicting when they will reach us. There is an emphasis on the perception of biological movement when the available visual information is impoverished. We also consider the role of the mirror neuron system in interpreting human movement.

Finally, we consider our ability (or failure!) to *detect* changes in objects within the visual environment over time. Unsurprisingly, attention importantly determines which aspects of the environment are consciously

detected. This issue provides a useful bridge between the areas of visual perception and attention (the subject of the next chapter).

DIRECT PERCEPTION

James Gibson (1950, 1966, 1979) put forward a radical approach to visual perception that was largely ignored at the time. Until approximately 40 years ago, it was assumed the main purpose of visual perception is to allow us to identify or recognise objects. This typically involves relating information extracted from the visual environment to our stored knowledge of objects (see Chapter 3). Gibson argued that this approach is limited – in evolutionary terms, vision developed so our ancestors could respond rapidly to the environment (e.g., hunting animals; escaping from danger).

Gibson (1979, p. 239) argued that perception involves "keeping in touch with the environment". This is sufficient for most purposes because the information provided by environmental stimuli is much richer than previously believed. We can relate Gibson's views to Milner and Goodale's (1995, 2008) vision-for-action system (see Chapter 2). According to both theoretical accounts, there is an intimate relationship between perception and action.

Gibson regarded his theoretical approach as ecological. He emphasised that perception facilitates interactions between the individual and their environment. Here is the essence of his direct theory of perception:

> When I assert that perception of the environment is direct, I mean that it is not mediated by *retinal* pictures, *neural* pictures, or *mental* pictures. *Direct perception* is the activity of getting information from the ambient array of light. I call this a process of *information pickup* that involves . . . looking around, getting around, and looking at things.
>
> (Gibson, 1979, p. 147)

We will briefly consider some of Gibson's theoretical assumptions:

- The pattern of light reaching the eye is an **optic array**. It contains all the visual information from the environment striking the eye.
- The optic array provides unambiguous or invariant information about the layout of objects. This information comes in many forms including optic flow patterns and affordances (see below) and texture gradients (discussed in Chapter 2).

Gibson produced training films in the Second World War describing how pilots handle taking off and landing. Of crucial importance is **optic flow** – the changes in the pattern of light reaching observers when they move or parts of the visual environment move. When pilots approach a landing strip, the point towards which they are moving (**focus of expansion**) appears motionless with the rest of the visual environment apparently moving away from that point (see Figure 4.1). The further away any part of the landing strip is from that point, the greater is its apparent speed of movement.

Wang et al. (2012) simulated the pattern of optic flow that would be experienced if individuals moved forwards in a stationary environment.
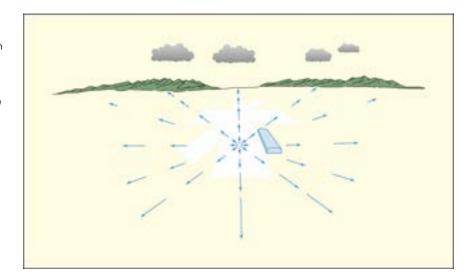
COMPANION WEBSITE

Case study:
Gibson's theory of direct perception affordances

Figure 4.1
The optic-flow field as a pilot comes in to land, with the focus of expansion in the middle.

From Gibson (1950). Wadsworth, a part of Cengage Learning, Inc. 2014 American Psychological Association. Reproduced with permission.

Their attention was attracted towards the focus of expansion, thus showing its psychological importance. (More is said later about optic flow and the focus of expansion.)

Gibson (1966, 1979) argued certain higher-order characteristics of the visual array (**invariants**) remain unaltered as observers move around their environment. Invariants (e.g., the focus of expansion) are important because they remain the same over different viewing angles. The focus of expansion is an invariant feature of the optic array.

Affordances

According to Gibson (1979), the potential uses of objects (their **affordances**) are directly perceivable. For example, a ladder "affords" ascent or descent. Gibson believed that "affordances are opportunities for action that exist in the environment and do not depend on the animal's mind . . . they do not cause behaviour but simply make it possible" (Withagen et al., 2012, p. 251). In Gibson (1979, p. 127), affordances are what the environment "*offers* the animal, what it *provides* or *furnishes*".

Evidence for the affordance of "climbability" of steps varying in height was reported by Di Stasi and Guardini (2007). The step height judged the most "climbable" was the one that would have involved the minimum expenditure of energy.

Gibson argued an object's affordances are perceived directly or automatically. In support, Pappas and Mack (2008) found images of objects presented below the level of conscious awareness nevertheless produced motor priming. For example, the image of a hammer caused activation in brain areas involved in preparing to use a hammer. Wilf et al. (2013) focused on the affordance of graspability with participants lifting their arms to perform a reach-like movement with graspable and non-graspable objects (see Figure 4.2). Muscle activity started faster for graspable than non-graspable objects suggesting that the affordance of graspability triggers rapid activity in the motor system.

KEY TERMS 🔑

Invariants
Properties of the optic array that remain constant even though other aspects vary; part of Gibson's theory.

Affordances
The potential uses of an object which Gibson claimed are perceived directly.

Gibson's approach to affordances is substantially oversimplified. For example, an apparently simple task such as cutting up a tomato involves selecting an appropriate tool, deciding on how to grasp and manipulate the tool, and monitoring movement execution (Osiurak & Badets, 2016). In other words, "People reason about physical object properties to solve everyday life activities" (Osiurak & Badets, 2016, p. 540). This is sharply different to Gibson's emphasis on the ease and immediacy of tool use.

When individuals observe a tool, Gibson assumed this provided them with *direct* access to knowledge about how to manipulate it and this manipulation knowledge gave access to the tool's functions. This assumption exaggerates the importance of manipulation knowledge. For example,

Garcea and Mahon (2012) found function judgements about tools were made faster than manipulation judgements, whereas Gibson's approach implies that manipulation judgements should have been faster.

Finally, Gibson argued stored knowledge is *not* required for individuals to make appropriate movements with respect to objects (e.g., tools). In fact, individuals often make extensive use of motor and function knowledge when dealing with objects (Osiurak & Badets, 2016). For example, making tea involves filling the kettle with water, boiling the water, finding some milk and so on. Foulsham (2015) discussed research showing there are only small individual differences in the pattern of eye fixations when people make tea. Such findings strongly imply they use stored information about the sequence of motor actions involved in tea-making.

Evaluation

What are the strengths of Gibson's ecological approach? First, "Gibson's realisation that *natural scenes* are the ecologically valid stimulus that should be used for the study of vision was of fundamental importance" (Bruce & Tadmor, 2015, p. 32).

Second, and related to the first point, Gibson disagreed with the previous emphasis on *static* observers looking at *static* visual displays. Foulsham and Kingstone (2017) compared the eye fixations of participants walking around a university campus with those of other participants viewing static pictures of the same scene. The eye fixations were significantly different: those engaged in walking focused more on features (e.g., the path) important for locomotion whereas those viewing static pictures focused centrally within each picture.

Third, Gibson was far ahead of his time. There is support for two visual systems (Milner & Goodale, 1995, 2008; see Chapter 2): a vision-for-perception system and a vision-for-action system. Before Gibson, the major emphasis was on the former. In contrast, he argued our perceptual system allows us to respond rapidly and accurately to environmental stimuli without using memory, which is a feature of the latter system.

What are the limitations of Gibson's approach? First, Gibson attempted to specify the visual information used to guide action but ignored many of the processes involved (see Chapters 2 and 3). For example, Gibson assumed the perception of invariants occurred almost "automatically", but it actually requires several complex processes.

Second, Gibson's argument that we do not need to assume the existence of internal representations (e.g., object knowledge) is flawed. The logic of Gibson's position is that: "There are invariants specifying a friend's face, a performance of Hamlet, or the sinking of the Titanic, and no knowledge of the friend, of the play, or of maritime history is required to perceive these things" (Bruce et al., 2003, p. 410). Evidence refuting Gibson's argument was reviewed by Foulsham (2015; discussed above).

Third, and related to the second point, Gibson de-emphasised the role of top-down processes (based on our knowledge and expectations) in visual perception. Such processes are especially important when the visual input is impoverished (see Chapter 3).

Fourth, Gibson's views on the effects of motion on perception were oversimplified. For example, when moving towards a goal, we use more information sources than Gibson assumed (discussed below).

VISUALLY GUIDED MOVEMENT

From an ecological perspective, it is important to understand how we move around the environment. For example, what information do we use when walking towards our current goal or target? We must ensure we are not hit by cars when crossing the road and when driving we must avoid hitting other cars. Playing tennis well involves predicting exactly *when* and *where* the ball will strike our racquet. The ways visual perception plays a crucial role in facilitating our locomotion and ensuring our safety are discussed in the next section.

Heading and steering

When we want to reach some goal (e.g., a gate at the end of a field), we use visual information to move directly towards it. Gibson (1950) emphasised the importance of optic flow (see Glossary; discussed on pp. 141–142). When we move forwards in a straight line, the point towards which we are moving (the focus of expansion) appears motionless. In contrast, the area around that point seems to expand.

Gibson (1950) proposed a hypothesis, according to which, if we are not moving directly towards our goal, we use the focus of expansion and optic flow to bring our heading (point of expansion) into alignment with our goal. This is known as the global radial outflow hypothesis.

Gibson's approach works well in principle when applied to an individual trying to move *straight* from A to B. However, matters are more complex when we cannot move directly to our goal (e.g., driving around a bend in the road; avoiding obstacles). Another complexity is that observers often make head and eye movements. In sum, the **retinal flow field** (changes in the pattern of light on the retina) is influenced by rotation in the retinal image produced by following a curved path and/or eye and head movements.

The above complexities mean it is often hard to use information from retinal flow to determine our direction of heading. It has often been claimed that a copy of motor commands (preprogramming) to move the eye and head (**efference copy**) is used by observers to compensate for the effects of eye and head movements on the retinal image. However, Feldman (2016) argued this approach is insufficient on its own because it de-emphasises the brain's active involvement in relating perception and action.

Findings: heading

Gibson emphasised the role of optic flow in allowing individuals to move directly towards their goal. Relevant information includes the focus of expansion (see Glossary) and the direction of radial motion (e.g., expansion within optic flow). Strong et al. (2017) obtained evidence indicating the importance of both factors and also established they depend on separate brain areas. More specifically, they used transcranial magnetic stimulation

(TMS; see Glossary) to disrupt key brain areas. TMS applied to area V3A impaired perception of the focus of expansion but not direction of radial motion, with the opposite pattern being obtained when TMS was applied to the motion area V5/MT+ (see Chapter 2).

As indicated above, eye and/or head movements make it harder to use optic flow effectively for heading. Bremmer et al. (2010) considered this issue in macaque monkeys presented with distorted visual flow fields simulating the combined effects of self-motion and an eye movement. Their key finding was that numerous cells in the medial superior temporal area successfully compensated for this distortion.

According to Gibson, a walker tries to make the focus of expansion coincide with the body moving straight ahead. If a walker wore prisms producing a 9° error in their perceived visual direction, the focus of expansion should be misaligned compared to their expectation. As a result, there should be a correction process, a prediction confirmed by Herlihey and Rushton (2012). Also as predicted, walkers denied access to information about retinal motion failed to show any correction.

Factors additional to the optic flow information emphasised by Gibson are also used when making heading judgements. This is unsurprising given the typical richness of the available environmental information. van den Berg and Brenner (1994) noted we only require one eye to use optic flow information. However, they discovered heading judgements were more accurate when observers used both eyes. Binocular disparity (see Glossary) in the two-eye condition provided useful additional information about the relative depths of objects. Cormack et al. (2017) introduced the notion of a binoptic flow field to describe the 3-D information available to observers (but de-emphasised by Gibson).

Gibson assumed optic-flow patterns generated by self-motion are of fundamental importance when we head towards a goal. However, motion is *not* essential for accurate perception of heading. The judgements of heading direction made by observers viewing two static photographs of a real-world scene in rapid succession were reasonably accurate in the absence of optic-flow information (Hahn et al., 2003). These findings can be explained in terms of retinal displacement – objects closer to the direction of heading show less retinal displacement as we move closer to the target.

Snyder and Bischof (2010) argued that information about the direction of heading is provided by two systems. One system uses *movement* information (e.g., optic flow) rapidly and fairly automatically (as proposed by Gibson). The other system uses *displacement* information more slowly and requires greater processing resources. It follows that performing a second task at the same time as making judgements about direction of heading should have little effect on those judgements if movement information is available. In contrast, a second task should impair heading judgements when only displacement information is available. The evidence supported both predictions.

Heading: future path

Wilkie and Wann (2006) argued judgements of heading (the direction in which someone is moving) are of little relevance if they are moving along a

curved path. With curved paths, path judge-ments (identifying future points along one's path) were much more accurate than heading judgements.

According to the above analysis, we might expect individuals (e.g., drivers) to fixate some point along their future path when it is curved. This is the future path strategy. In contrast, Land and Lee (1994) argued (with supportive evidence) that drivers approaching a bend focus on the **tangent point** – the point on the inside edge of the road at which its direction appears to reverse (see Figure 4.3).

The tangent point has two potential advantages. First, it is easy to identify and track. Second, road curvature can easily be worked out by considering the angle between the direction of heading and the tangent point. Kandil et al. (2009) found most drivers nego-tiating 270° bends at a motorway junction fixated the tangent point much more often than the future path (75% vs 14%, respectively).
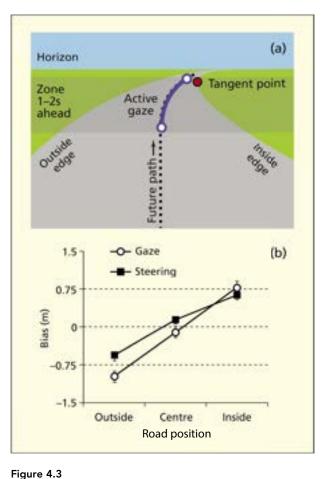
Other research suggests the tangent point is less important. For example, Itkonen et al. (2015) instructed drivers to "drive as they normally would" or "look at the tangent point". Eye movements differed markedly in the two conditions – drivers were much more likely to fixate points along the future path in the former condition.

How can we interpret the above appar-ently inconsistent findings? Lappi et al. (2013) hypothesised drivers often fixate the tangent point when approaching and entering a bend but fixate the future path further into the bend. They argued the tangent point provides relatively precise informa-tion and so drivers use it when uncertainty about the precise nature of the curve or bend is maximal (i.e., when approaching and entering it).

Lappi et al. (2013) obtained supporting evidence for the above hypoth-esis. Drivers' fixations while driving along a lengthy curve formed by the slip road to a motorway were predominantly on the path ahead rather than the tangent point after the first few seconds (see short clips of drivers' eye movements while performing this task at 10.1371/journal.pone.0068326).

The evidence discussed so far does not rule out optic flow as a factor influencing drivers' steering. Mole et al. (2016) manipulated optic-flow speed in a simulated driving situation. This produced steering errors (understeering or oversteering) when going around bends even when full information about road edges was available. Thus, optic flow influenced driving performance.

Figure 4.3

The visual features of a road viewed in perspective. The tangent point is marked by the filled circle on the inside edge of the road, and the desired future path is shown by the dotted line. According to the future-path theory, drivers should gaze along the line marked "active gaze".

From Wilkie et al. (2010). Reprinted with permission from Springer-Verlag.

KEY TERM

Tangent point
From a driver's perspective, the point on a road at which the direction of its inside edge appears to reverse.

IN THE REAL WORLD: ON-ROAD DRIVING

Much research on drivers' gaze patterns lacks ecological validity (see Glossary). Drivers are typically in a simulator and the environment through which they drive is somewhat oversimplified. Accordingly, Lappi et al. (2017) studied the gaze patterns of a 43-year-old male driving school instructor driving on a rural road in Finland. His eye movements revealed a more complex picture than most previous research.

What did Lappi et al. (2017) discover? Here are four major findings:

(1) The driver's gaze shifted very frequently from one feature of the visual environment to another and he made many head movements.
(2) The driver's gaze was predominantly on the far road (see Figure 4.4). This preview of the road ahead allowed him to make use of anticipatory control.
(3) In bends, the driver's gaze was mostly within the far road "triangle" formed by the tangent point (TP), the lane edge opposite the TP and the occlusion point (OP; the point where the road disappears from view). In general terms, the OP is used to anticipate the road ahead whereas the TP is used for more immediate compensatory steering control.
(4) The driver fixated specific targets (e.g., traffic signs; other road users) very rapidly, suggesting his peripheral vision was very efficient.

(A)   (B)

Figure 4.4
The far road "triangle" in (A) a left turn and (B) a right turn.
From Lappi et al. (2017).

In sum, drivers' gaze patterns are more complex than implied by previous research. Drivers do *not* constantly fixate any given feature (e.g., tangent point) *passively*. Instead, they "sample visual information as needed, leading to input that is intermittent, and determined by the active observer . . . rather than imposed by the environment" (Lappi et al., 2017, p. 11). Drivers' eye movements are determined in part by control mechanisms (e.g., path planning) (Lappi & Mole, 2018). These mechanisms are responsive to drivers' goals. For example, professional racing drivers have the goal of driving as fast as possible whereas many ordinary drivers have the goal of driving safely.

Evaluation

Gibson's views concerning the importance of optic-flow information have deservedly been very influential. Such information is especially useful when individuals can move *directly* towards their goal rather than following a curved or indirect path. Indeed, the evidence suggests optic flow is often

the dominant source of information determining judgements of heading direction. Drivers going around bends use optic-flow information. They also make some use of the tangent point. This is a relatively simple feature of the visual environment and its use by drivers is in the spirit of Gibson's perspective.

What are the limitations of Gibson's approach and other related approaches?

(1) Individuals moving directly towards a target use several kinds of information (e.g., binocular disparity; retinal displacement) ignored by Gibson.

(2) The tangent point is used infrequently when individuals move along a curved path: they more often fixate points lying along the future path.

(3) Drivers going around bends use a greater variety of information sources than implied by Gibson's approach. Of most importance, drivers' eye movements are strongly influenced by active, top-down processes (e.g., motor control) not included within Gibson's theorising. More specifically, drivers' eye movements depend on their current driving goals as well as the environmental conditions.

(4) Research and theorising have de-emphasised meta-cognition (beliefs about one's own performance). Mole and Lappi (2018) found drivers often made inaccurate meta-cognitive judgements of their own driving performance (e.g., they tended to exaggerate the importance of driving speed in determining performance). Such inaccurate judgements probably often lead to impaired driving performance.

Time to contact

In everyday life, we often need to predict the moment of contact between us and some object. These situations include ones where we are moving towards an object (e.g., a wall) and those in which an object (e.g., a ball) is approaching us. We might work out the time to contact by dividing our estimate of the object's distance by our estimate of its speed. However, this would be complex and error-prone because information about speed and distance is not *directly* available.

Lee (1976, 2009) argued that there is a simpler way to work out the time to contact or collision. If we approach it (or it approaches us) at constant velocity, we can use tau. Tau is defined as the size of an object's retinal image divided by its rate of expansion. The faster the rate of expansion, the less time there is to contact.

When driving, the rate of decline of tau over time (tau-dot) indicates whether there is sufficient braking time to stop before contact or collision. Lee (1976) argued drivers brake to hold constant the rate of change of tau. This tau-dot hypothesis is consistent with Gibson's approach because it assumes tau-dot is an invariant available to observers from optic flow.

Lee's theoretical approach has been highly influential. However, his emphasis on tau has limited applicability in various ways (Tresilian, 1999). First, tau ignores acceleration in object velocity. Second, tau only provides information about the time to contact or collision with the eyes. Thus, drivers might find the front of their car smashed in if they relied solely on

tau! Third, tau is accurate only when applied to spherically symmetrical objects: do not rely on it when catching a rugby ball!

Harrison et al. (2016) argued that people's behaviour is often influenced by factors other than their estimate of the time to contact. For example, consider someone deciding whether to cross a road when there is an approaching car. Their decision is often influenced by judgements of their physical mobility and their personality (e.g., cautious or impetuous) (see p. 151).

Findings

According to Lee's (1976) theory, observers can often judge time to contact accurately based on using tau relatively "automatically". If so, observers' time-to-contact judgements might not be impaired if they performed a cognitively demanding task while observing an object's movement. Baurès et al. (2018) obtained support for this prediction. Indeed, time-to-contact judgements were *more* accurate when observers performed a secondary task, perhaps because this made it less likely they would attend to potentially misleading information (e.g., expectations about an object's movements).

According to Lee (1976), judgements of the time to contact when catching a ball should depend crucially on the rate of expansion of the ball's retinal image. Savelsbergh et al. (1993) used a deflating ball having a significantly slower rate of expansion than an ordinary ball. The prediction was that peak grasp closure should occur *later* to the deflating ball. This prediction was confirmed. However, the actual slowing was much less than predicted (30 ms vs 230 ms). Participants minimised the distorting effects of manipulating the rate of expansion by using additional sources of information (e.g., depth cues).

Hosking and Crassini (2010) had participants judge time to contact for familiar objects (tennis ball and football) presented in their standard size or with their sizes reversed. They also used unfamiliar black spheres. Contrary to Lee's hypothesis, time-to-contact judgments were influenced by familiar size (especially when the object was a very large tennis ball) leading participants to overestimate time to contact (see Figure 4.5).

Tau is available in monocular vision. However, observers often make use of information available in binocular vision, especially binocular disparity (see Chapter 2). Fath et al. (2018) discussed research showing binocular information sometimes provides more accurate judgements than tau of time to contact (e.g., when viewing small objects or rotating non-spherical objects).
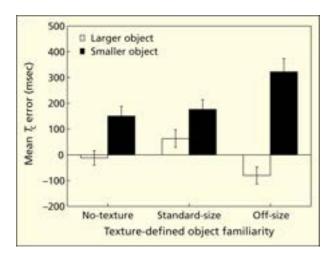

Figure 4.5

Errors in time-to-contact judgements for the smaller and the larger object as a function of whether they were presented in their standard size, the reverse size (off-size) or lacking texture (no-texture). Positive values indicate that responses were made too late and negative values that they were made too early.

From Hosking and Crassini (2010). With kind permission from Springer Science+Business Media.

In their own research, Fath et al. (2018) assessed accuracy of time-to-contact judgements when observers viewed fast- or slow-moving objects. They used three conditions varying in the amount of information available to observers: (1) monocular flow information only (permitting assessment of tau); (2) binocular disparity information only; (3) all sources of information available. Fath et al. predicted that binocular disparity information would be less likely to be used with fast-moving objects than slow-moving ones because it is relatively time-consuming to calculate changes in binocular disparity over time.

What did Fath et al. (2018) find? First, with fast objects, time-to-contact judgements were more accurate with monocular flow information only than with binocular disparity information only. Second, with slow objects, the opposite findings were obtained. Third, accuracy of time-to-contact judgements when all sources of information were available were comparable to accuracy in the better of the single-source conditions with fast and with slow objects.

DeLucia (2013) found observers mistakenly predicted a large approaching object would reach them sooner than a closer small approaching object: the size-arrival effect. This effect occurred because observers attached more importance to relative size than tau.

We turn now to research on drivers' braking decisions. Lee's (1976) notion that drivers brake to hold constant the rate of change of tau was tested by Yilmaz and Warren (1995). They told participants to stop at a stop sign in a simulated driving task. As predicted, there was generally a linear reduction in tau during braking. However, some participants showed large rather than gradual changes in tau shortly before stopping.

Tijtgat et al. (2008) found individual differences in stereo vision influenced drivers' braking behaviour to avoid a collision. Drivers with weak stereo vision started braking earlier than those with normal stereo vision and their peak deceleration also occurred earlier. Those with weak stereo vision found it harder to calculate distances causing them to underestimate the time to contact. Thus, deciding when to brake does not depend only on tau or tau-dot.

Harrison et al. (2016) argued that Lee's (1976) theoretical approach is limited in two important ways when applied to drivers' braking behaviour. First, it ignores physical limitations in the real world. For example, tau-dot specifies to a driver the deceleration during braking required to avoid collision. However, this strategy will not work if the driver's braking system makes the required deceleration unachievable.

Second, individuals differ in the emphasis they place on minimisation of costs (e.g., preferred safety margin). According to Harrison et al., these limitations suggest drivers' braking behaviour is influenced by their sensitivity to relevant affordances (possibilities for action) such as their knowledge of the dynamics of the braking system in their car.

Evaluation

The notion that tau is used to make time-to-contact judgements is simple and elegant. There is much evidence that such judgements are often strongly influenced by tau. Even when competing factors affect time-to-contact

judgements, tau often has the greatest influence on those judgements. Tau is also often used when drivers make decisions about when to brake.

What are the limitations of theory and research in this area? First, time-to-contact judgements are typically more influenced by tau or tau-dot in relatively uncluttered laboratory environments than naturalistic conditions (Land, 2009). Second, tau is not the only factor determining time-to-contact judgements. As Land (2009, p. 853) pointed out, "The brain will accept all valid cues in the performance of an action, and weight them according to their current reliability." These cues can include object familiarity, binocular disparity and relative size. It clearly makes sense to use all the available information in this way.

Third, the tau hypothesis ignores the *emotional* value of the approaching object. Time-to-contact judgements are shorter for threatening pictures than neutral ones (Brendel et al., 2012). This makes evolutionary sense – it could be fatal to overestimate how long a very threatening object (e.g., a lion) will take to reach you!

Fourth, braking behaviour involves factors additional to tau and tau-dot. For example, there are individual differences in preferred safety margin. Rock et al. (2006) identified an alternative braking strategy in a real-world driving task in which drivers directly estimated the constant ideal deceleration required to stop at a given point.

VISUALLY GUIDED ACTION: CONTEMPORARY APPROACHES

The previous section focused mainly on the issue of how we use visual information when moving through the environment. Here we consider similar issues but the emphasis shifts towards processes involved in successful goal-directed action towards objects. For example, how do we reach for a cup of coffee? This issue was addressed by Milner and Goodale (1995, 2008) in their perception-action model (see Chapter 2). Contemporary approaches that have developed and extended the perception-action model are discussed below.

Role of planning: planning-control model

Glover (2004) proposed a planning-control model of goal-directed action towards objects. According to this model, we initially use a planning system followed by a control system but the two systems often overlap in time. Here are the main features of the two systems:

(1)   *Planning system*
   - It is used mostly *before* the initiation of movement.
   - It selects an appropriate target (e.g., cup of coffee), decides how it should be grasped and works out the timing of the movement.
   - It is influenced by factors such as the individual's goals, the nature of the target object, the visual context and various cognitive processes.
   - It is relatively slow because it uses much information and is influenced by conscious processes.

(2) *Control system*
- It is used during the carrying out of a movement.
- It ensures movements are accurate, making adjustments, if necessary, based on visual feedback. Efference copy (see Glossary) is used to compare actual with desired movement. **Proprioception** is also involved.
- It is influenced by the target object's spatial characteristics (e.g., size; shape; orientation) but not by the surrounding context.
- It is fairly fast because it uses little information and is not susceptible to conscious influence.

According to the planning-control model, most errors in human action stem from the planning system. In contrast, the control system typically ensures actions are accurate and achieve their goal. Many visual illusions occur because of the influence of visual context. Since information about visual context is used only by the planning system, responses to visual illusions should typically be inaccurate if they depend on the planning system but accurate if they depend on the control system.

There are similarities between the planning-control model and Milner and Goodale's perception-action model. However, Glover (2004) focused more on the processing *changes* occurring during action performance.

Findings

Glover et al. (2012) compared the brain areas involved in planning and control using a planning condition (prepare to reach and grasp an object but remain still) and a control condition (reach out immediately for the object). There was practically no overlap in the brain areas associated with planning and control. This finding supports the model's assumption that planning and control processes are separate.

According to the planning-control model, various factors (e.g., semantic properties of the visual scene) influence the planning process associated with goal-directed movements but not the subsequent control process. This prediction was tested by Namdar et al. (2014). Participants grasped an object in front of them using their thumb and index finger. The object had a task-irrelevant digit (1, 2, 8 or 9) on it. As predicted, numerically larger digits led to larger grip apertures during the first half of the movement trajectory but not the second half (involving the control process).

According to Glover (2004), action planning involves conscious processing followed by rapid non-conscious processing during action control. These theoretical assumptions can be tested by requiring participants to carry out a second task while performing an action towards an object. According to the model, this second task should disrupt planning but not control. However, Hesse et al. (2012) found a second task disrupted planning *and* control when participants made grasping movements towards objects. Thus, planning and control can both require attentional resources.

According to the model, visual illusions occur because misleading visual context influences the initial planning system rather than the later

control system. Roberts et al. (2013) required participants to make rapid reaching movements to a Müller-Lyer figure. Vision was available only during the first 200 ms of movement or the last 200 ms. The findings were opposite to those predicted theoretically – performance was more accurate with early vision than late vision.

Elliott et al. (2017) explained the above findings with their multiple process model. According to this model, performance was good when early vision was available because of a control system known as impulse control. Impulse control "entails an early, and continuing, comparison of expected sensory consequences to perceived sensory consequences to regulate limb direction and velocity during the distance-covering phase of the movement" (p. 108).

Evaluation

Glover's (2004) planning-control model has proved successful in various ways. First, it successfully developed the common assumption that motor movements towards an object involve successive planning and control processes. Second, the assumption cognitive processes are important in action planning is correct. Third, there is evidence (e.g., Glover et al., 2012) that separate brain areas are involved in planning and control.

What are the model's limitations? First, the planning system involves several very different processes: "goal determination; target identification and selection; analysis of object affordances [potential object uses]; timing; and computation of the metrical properties of the target such as its size, shape, orientation and position relative to the body" (Glover et al., 2012, p. 909). This diversity sheds doubt on the assumption there is a *single* planning system.

Second, the model argues control occurs late during object-directed movements and is influenced by visual feedback. However, there appears to be a second control process (called impulse control by Elliott et al., 2017) operating throughout the movement trajectory and not influenced by visual feedback.

Third, and related to the second point, the model presents an oversimplified picture of the processes involved in goal-directed action. More specifically, the processing involved in producing goal-directed movements is far more complex than implied by the notion of a planning process followed by a control process. For example, planning and control processes are often so intermixed that "the distinction between movement planning and movement control is blurred" (Gallivan et al., 2018, p. 519).

Fourth, complex decision-making processes are often involved when individuals plan goal-directed actions in the real world. For example, when planning, tennis players players must often decide between a simple shot minimising energy expenditure and risk or injury or a more ambitious shot that might immediately win the current point (Gallivan et al., 2018).

Fifth, the model is designed to account for planning and control processes when only *one* object is present or of interest. In contrast, visual scenes in everyday life are often far more complex and contain several objects of potential relevance (see below).

Role of planning: changing action plans

We all have considerable experience of changing, modifying and abandoning action plans with respect to objects in the environment. How do we resolve competition among action plans? According to Song (2017, p. 1), "Critical is the existence of parallel motor planning processes, which allow efficient and timely changes."

What evidence indicates we often process information about several different potential actions simultaneously? Suppose participants are given the task of reaching rapidly towards a target in the presence of distractors (Song & Nakayama, 2008). On some trials, their reach is initially directed towards the target. On other trials, their initial reach is directed towards a distractor but this is corrected in mid-flight producing a strongly curved trajectory. Song and Nakayama's key finding was that corrective movements occurred very rapidly following the onset of the initial movement. This finding strongly implies that the corrective movement had been planned *prior* to execution of the initial incorrect movement.

Song (2017) discussed several other studies where similar findings were obtained. He concluded, "The sensori-motor system generates multiple competing plans in parallel before actions are initiated . . . this concurrent processing enables us to efficiently resolve competition and select one appropriate action rapidly" (p. 6).

Brain pathways

In their perception-action model, Milner and Goodale (1995, 2008) distinguished between a ventral stream or pathway and a dorsal stream or pathway (see Chapter 2). In approximate terms, the ventral stream is involved in object perception whereas the dorsal stream "is generally considered to mediate the visual guidance of action, primarily in real time" (Milner, 2017, p. 1297).

Interactive feature:
Primal Pictures'
3D atlas of the brain

Much recent research has indicated that the above theoretical account is oversimplified (see Chapter 2). Of central importance is the accumulating evidence that there are actually *two* somewhat separate dorsal streams (Osiurak et al., 2017; Sakreida et al., 2016):
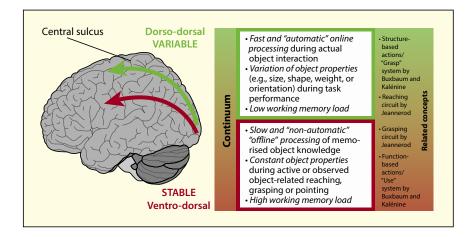
(1) The dorso-dorsal stream: processing in this stream relates to the online control of action and is *hand*-centred; it has been described as the "grasp" system (Binkofski & Buxbaum, 2013).

(2) The ventro-dorsal stream: processing in this stream is offline and relies on memorised knowledge of objects and tools and is *object*-centred; it has been described as the "use" system (Binkofski & Buxbaum, 2013).

Sakreida et al. (2016) identified several other differences between these two streams (see Figure 4.6). In essence, object processing within the dorso-dorsal stream is *variable* because it is determined by the immediately accessible properties of an object (e.g., its size and shape). Such processing is fast and "automatic". In contrast, processing within the ventro-dorsal stream is *stable* because it is determined by memorised object knowledge.

Figure 4.6
The dorso-dorsal and ventro-dorsal streams showing their brain locations and forms of processing.

From Sakreida et al. (2016). Reprinted with permission of Elsevier.

Such processing is slow and more cognitively demanding than processing within the dorso-dorsal stream.

Findings

Considerable neuroimaging evidence supports the proposed distinction between two dorsal streams. Martin et al. (2018, p. 3755) reviewed research indicating the dorso-dorsal stream "traverses from visual area V3a through V6 toward the superior parietal lobule, and . . . reaches the dorsal premotor cortex". In contrast, the ventro-dorsal stream "encompasses higher-order visual areas like MT/V5+, the inferior parietal lobule . . . as well as the ventral premotor cortex and inferior frontal gyrus" (p. 3755). Sakreida et al. (2016) conducted a meta-analytic review based on 71 neuroimaging studies and obtained similar findings.

Evidence from brain-damaged patients is also supportive of the distinction between two dorsal streams. First, we consider patients with damage to the ventro-dorsal stream. Much research has focused on **limb apraxia**, a disorder where patients often fail to make precise goal-directed actions in spite of possessing the physical ability to perform those actions (Pellicano et al., 2017). More specifically, "Reaching and grasping actions in LA [limb apraxia] are normal when vision of the limb and target is available, but typically degrade when they must be performed 'off-line', as when subjects are blindfolded prior to movement execution" (Binkovski & Buxbaum, 2013, p. 5). This pattern of findings is expected if the dorso-dorsal stream is intact in patients with limb apraxia.

Second, we consider patients with damage to the dorso-dorsal stream. Much research here has focused on optic ataxia (see Glossary). As predicted, patients with optic ataxia have impaired online motor control and so exhibit inaccurate reaching towards (and grasping of) objects.

Evaluation

Neuroimaging research has provided convincing evidence for the existence of two dorsal processing streams. The distinction between dorso-dorsal and ventro-dorsal streams has also been supported by studies on brain-damaged

KEY TERM 🔑

Limb apraxia
A condition caused by brain damage in which individuals have impaired ability to make skilled goal-directed movements towards objects even though they possess the physical ability to perform them.

patients. More specifically, there is some evidence for a double dissociation (see Glossary) between the impairments exhibited by patients with limb apraxia and optic ataxia.

What are the limitations of research in this area? First, the ventral stream (strongly involved in object recognition) is also important in visually guided action (Osiurak et al., 2017). However, precisely how this stream interacts with the dorso-dorsal and ventro-dorsal streams is unclear. Second, there is some *overlap* in the brain between the dorso-dorsal and ventro-dorsal streams and so it is important not to exaggerate their independence. Third, there is a lack of consensus concerning the precise functions of the two dorsal streams (see Osiurak et al., 2017, and Sakreida et al., 2016).

PERCEPTION OF HUMAN MOTION

We are very good at interpreting other people's movements. We can decide very rapidly whether someone is walking, running or limping. Our initial focus is on two key issues. First, how successfully can we interpret human motion with very limited visual information? Second, do the processes involved in perception of *human* motion differ from those involved in perception of motion *in general*? If the answer to this question is positive, we also need to consider *why* the perception of human motion is special.

As indicated already, our focus is mostly on the perception of human motion. However, there are many similarities between the perception of human and animal motion, and we will sometimes use the term "biological motion" to refer generally to the perception of animal motion.

Finally, we discuss an important theoretical approach based on the notion that the same brain system or network is involved in perceiving and understanding human actions and in performing those same actions.

Perceiving human motion

Suppose you were presented with point-light displays, as was done initially by Johansson (1973). Actors were dressed entirely in black with lights attached to their joints (e.g., wrists; knees; ankles). They were filmed moving around a darkened room so only the lights were visible to observers watching the film (see Figure 4.7 and "Johansson Motion Perception Part 1" on YouTube).What do you think you would perceive in those circumstances?
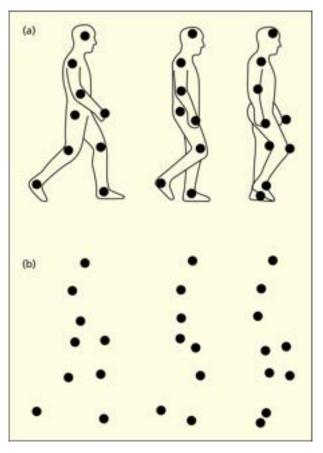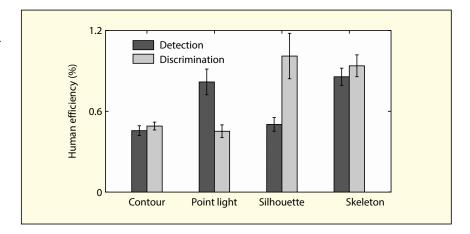

Figure 4.7

Point-light sequences (a) with the walker visible and (b) with the walker not visible.

Shiffrar and Thomas (2013). With permission of the authors.

Figure 4.8

Human detection and discrimination efficiency for human walkers presented in contour, point lights, silhouette and skeleton.

From Lu et al. (2017).

In fact, Johansson found observers perceived the moving person accurately with only six lights and a short segment of film.

In subsequent research, Johansson et al. (1980) found observers perceived human motion with no apparent difficulty when viewing a point-light display for only one-fifth of a second! Ruffieux et al. (2016) studied a patient, BC, who was cortically blind but had a residual ability to process motion. When presented with two point-light displays (one of a human and one of an animal) at the same time, he generally correctly identified the human.

The above findings imply we are very efficient at processing impoverished point-light displays. However, Lu et al. (2017) reported some contrary evidence. Observers were given two tasks: (1) detecting the presence of a human walker; (2) discriminating whether a human walker was walking leftward or rightward. The walker was presented in point lights, contour, silhouette or as a skeleton. *Detection* performance was relatively good for the point-light display but *discrimination* performance was not (see Figure 4.8). Performance was high with the skeleton display because it provided detailed information about the connections between joints.

Top-down or bottom-up processes?

Johansson (1975) argued the ability to perceive biological motion is innate, describing the processes involved as "spontaneous" and "automatic". Support was reported by Simion et al. (2008) in a study on newborns (1–3 days). These newborns preferred to look at a display showing biological motion more than one that did not. Remarkably, Simion et al. used point-light displays of chickens of which the newborns had no previous experience. These findings suggest the perception of biological motion involves basic bottom-up processes.

Evidence that learning plays a role was reported by Pinto (2006). Three-month-olds were equally sensitive to motion in point-light humans, cats and spiders. In contrast, 5-month-olds were more sensitive to displays of human motion. Thus, the infant visual system becomes increasingly specialised for perceiving human motion.

If the detection of biological motion were "automatic", it would be relatively unaffected by attention. However, in a review Thompson and Parasuraman (2012) concluded attention is required, especially when the available visual information is ambiguous or competing information is present.

Mayer et al. (2015) presented circular arrays of between two and eight video clips. In one condition, observers decided rapidly whether any clip showed human motion; in another condition, they decided whether any clips showed machine motion. There were two key findings. First, detection times increased with array size for both human and machine motion, suggesting attention is required to detect both types of motion. Second, the effects of array size on detection times were much greater for machine motion. Thus, searching is more efficient for human than machine motion suggesting human motion perception may be special (see below).

Is human motion perception special?

Much evidence indicates we are better at detecting human motion than motion in other species (Shiffrar & Thomas, 2013). Cohen (2002) assessed observers' sensitivity to human, dog and seal motion using point-light displays. Performance was best with human motion and worst with seal motion. Of importance, the same pattern of performance was found in seal trainers and dog trainers. Thus, the key factor is *not* simply visual experience; instead, we are more sensitive to observed motions resembling our own repertoire of actions.

We can also consider whether human motion perception is special by considering the brain. There has been an increasing recognition that many brain areas are involved in biological motion processing (see Figure 4.9). The pathway from the fusiform gyrus (FFG) to the superior temporal sulcus (STS) is of particular importance, as are top-down processes from the insula (INS), the STS and the inferior frontal gyrus (IFG).

Much research indicates the central importance of the superior temporal sulcus. Grossman et al. (2005) applied repetitive transcranial magnetic stimulation (rTMS; see Glossary) to that area to disrupt processing. This caused a substantial reduction in observers' sensitivity to biological motion. Gilaie-Dotan et al. (2013) found grey matter volume in the superior temporal sulcus correlated positively with the detection of biological (but not non-biological) motion.

Evidence from brain-damaged patients indicates that perceiving biological motion involves different processes from those involved in perceiving object motion generally. Vaina et al. (1990) studied a patient, AF, with damage to the posterior visual pathways. He performed poorly on basic motion
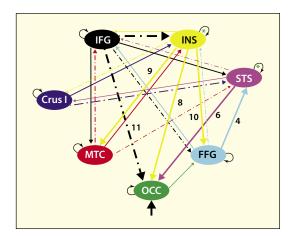

Figure 4.9

Brain areas involved in biological motion processing (STS = superior temporal sulcus; IFG = inferior frontal gyrus; INS = insula; Crus 1 = left lateral cerebellar lobule; MTC = middle temporal cortex; OCC = early visual cortex; FFG = fusiform gyrus).

From Sokolov et al. (2018).

tasks but was reasonably good at detecting biological motion from point-light displays. In contrast, Saygin (2007) found in stroke patients with damage in the temporal and premotor frontal areas that their perception of biological motion was more impaired than non-biological motion.

Why is biological motion perception special?

We could explain the special nature of biological motion perception in three ways (Shiffrar & Thomas, 2013). First, biological motion is the only type of motion humans can *produce* as well as perceive. Second, most people spend more time perceiving and trying to understand other people's motion than any other form of visual motion. Third, other people's movements provide a rich source of social and emotional information.

We start with the first reason (discussed further on pp. 161–162). The relevance of motor skills to the perception of biological motion was shown by Kloeters et al. (2017). Patients with Parkinson's disease (which impairs movement execution) had significantly inferior perception of human movement with point-light displays compared to healthy controls. More dramatically, paraplegics with severe spinal injury were almost *three* times less sensitive than healthy controls to human movement in point-light displays.

We must not exaggerate the importance of motor involvement in biological motion perception. A man, DC, born without upper limbs, identified manual actions shown in videos and photographs as well as healthy controls (Vannuscorps et al., 2013). Motor skills may be most important in biological motion perception when the visual information presented is sparse or ambiguous (e.g., as with point-light displays).

Jacobs et al. (2004) obtained support for the second reason listed above. Observers' ability to identify walkers from point-light displays was much better when the walker was observed for 20 hours a week rather than 5 hours. In our everyday lives, we often recognise individuals in motion by integrating information from biological motion with information from the face and the voice within the superior temporal sulcus (Yovel & O'Toole, 2016). Successful integration of these different information sources clearly depends on learning and experience.

We turn now to the third reason mentioned earlier. Charlie Chaplin showed convincingly that bodily movements can convey social and emotional information. Atkinson et al. (2004) found observers performed well at identifying emotions from point-light displays (especially for fear, sadness and happiness). Part of the explanation for these findings is that angry individuals walk especially fast whereas fearful or sad ones walk very slowly (Barliya et al., 2013).

We can explore the role of social factors in biological motion detection by studying adults with autism spectrum disorder who have severely impaired social interaction skills. The findings are somewhat inconsistent However, adults with autism spectrum disorder generally have a reasonably intact ability to detect human motion in point-light displays but exhibit impaired emotion processing in such displays (see Bakroon & Lakshminarayanan, 2018 for a review).

Mirror neuron system

Research on monkeys in the 1990s transformed our understanding of biological motion. Gallese et al. (1996) assessed monkeys' brain activity while they *performed* a given action and while they *observed* another monkey perform the same action. They found 17% of neurons in area F5 of the premotor cortex were activated in both conditions. Such findings led theorists to propose a **mirror neuron system** consisting of neurons activated during both observation *and* performance of actions (see Keysers et al., 2018, for a review).

There have been numerous attempts to identify a mirror neuron system in humans. Our current understanding of brain areas associated with the mirror neuron system is shown in Figure 4.10. Note that the mirror neuron system consists of an integrated network rather than separate brain areas (Keysers et al., 2018).

Most research is limited because it shows only that the same brain *areas* are involved in action perception and production. Perry et al. (2018) used more precise methods to reveal a more complex picture within areas assumed to form part of the mirror neuron system. Some small areas were activated during both observing actions and imitating them, thus providing evidence for a human neuron system. However, other adjacent areas were activated only during observing or action imitation.

More convincing evidence for a human mirror neuron system was reported by de la Rosa et al. (2016). They focused on activation in parts of the inferior frontal gyrus (BA44/45) corresponding to area F5 in monkeys. Their key finding was that 52 voxels (see Glossary) within BA44/45 responded to both action perception and action production.

Before proceeding, we should note the term "mirror neuron system" is somewhat misleading because mirror neurons do *not* provide us with an exact motoric coding of observed actions. As Williams (2013, p. 2962) wittily remarked, "If only this was the case! I could become a Olympic ice-skater or a concert pianist!"

Findings

We have seen that neuroimaging studies have indicated that the mirror neuron system is activated during motor perception and action. Such evidence is *correlational*, and so does not demonstrate that the mirror neuron system is necessary for motor perception and action understanding.

More direct evidence comes from research on brain-damaged patients. Binder et al. (2017) studied left-hemisphere stroke patients with **apraxia** (impaired ability to perform planned actions) having damage within the mirror neuron system (e.g., inferior frontal gyrus). These patients had comparable deficits in action imitation, action recognition and action
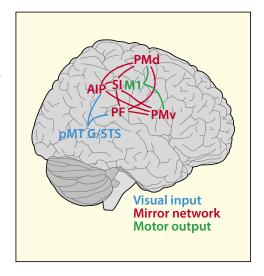
Figure 4.10

The main brain areas associated with the mirror neuron system (MNS) plus their interconnections (red). Areas involved in visual input (blue; pMTG = posterior mid-temporal gyrus; STS = superior temporal gyrus) and motor output (green; M1 = primary motor cortex) are also shown. AIP = anterior intraparietal areas; PF = area within the parietal lobe; PMv and PMd = ventral and dorsal premotor cortex; SI = primary somato-sensory cortices.

From Keysers et al. (2018). Reprinted with permission of Elsevier.

KEY TERM

Mirror neuron system
Neurons that respond to actions whether performed by oneself or someone else; it is claimed these neurons assist in imitating (and understanding) the actions of others.

Apraxia
A condition caused by
brain damage in which
there is greatly reduced
ability to perform
purposeful or planned
bodily movements in
spite of the absence of
muscular damage.

comprehension. The co-existence of these deficits was precisely as pre-dicted. Another predicted finding was that left-hemisphere stroke patients *without* apraxia had less brain damage in core regions of the mirror neuron system than those with apraxia.

Another approach to demonstrating the causal role of the mirror neuron system is to use experimental techniques such as transcranial direct current stimulation (tDCS; see Glossary). Avenanti et al. (2018) assessed observers' ability to predict which object would be grasped after seeing the start of a reaching movement. Task performance was *enhanced* when anodal tDCS was used to facilitate neural activity within the mirror neuron system, whereas it was *impaired* when cathodal tDCS was used to inhibit such neural activity.

Findings: functions of the mirror neuron system

What are the functions of the mirror neuron system? It has often been assumed mirror neurons play a role in working out *why* someone else is performing certain actions as well as deciding *what* those actions are. For example, Eagle et al. (2007, p. 131) claimed the mirror neuron system is involved in "the automatic, unconscious, and non-inferential simulation in the observer of the actions, emotions, and sensations carried out and expressed by the observed".

Rizzolatti and Sinigaglia (2016) argued that full understanding of another person's actions requires a multi-level process. The first level involves identifying the outcome of the observed action and the emotion being displayed by the other person. This is followed by the observer rep-resenting the other person's desires, beliefs and intentions. The mirror neuron system is primarily involved at the first level but may provide an input to subsequent processes.

Lingnau and Petris (2013) argued that understanding another person's actions often requires complex *cognitive* processes as well as simpler pro-cesses within the mirror neuron system. Observers saw point-light displays of human actions and some were asked to identify the goal of each action. Areas within the prefrontal cortex (associated with high-level cognitive pro-cesses) were more activated when goal identification was required. These findings can be explained within the context of Rizzolatti and Sinigaglia's (2016) approach discussed above.

Wurm et al. (2016) distinguished between two forms of motion per-ception and understanding. They used the example of observers under-standing that someone is opening a box. If they have a general or abstract understanding of this action, their understanding should *generalise* to other boxes and other ways of opening a box. In contrast, if they only have a specific or concrete understanding of the action, their understand-ing will *not* generalise. Wurm et al. (2016) found specific or concrete action understanding could occur within the mirror neuron system. However, more general or abstract understanding involved high-level perceptual regions (e.g., the lateral parieto-temporal cortex) *outside* the mirror neuron system.

In sum, the mirror neuron system is of central importance with respect to some (but not all) aspects of action understanding. More specifically,

additional (more "cognitive") brain areas are required if action under-standing is complex (Lingnau & Petris, 2013) or involves generalising from past experience (Wurm et al., 2016). It is also likely that imitating someone else's actions often involves processes (e.g., person-perception processes) additional to those directly involving the mirror neuron system (Ramsey, 2018).

Overall evaluation

Several important research findings have been obtained. First, we have an impressive ability to perceive human or biological motion even with very limited visual input. Second, the brain areas involved in human motion perception differ somewhat from those involved in perceiving motion in general. Third, perception of human motion is special because it is the only type of motion we can both perceive *and* produce. Fourth, a mirror neuron system allows us to imitate and understand other people's movements. Fifth, the core brain network of the mirror neuron system has been identi-fied. Its causal role has been established through studies on brain-damaged patients and research using techniques such as transcranial direct current stimulation.

What are the limitations of research in this area? First, much remains unclear about interactions of bottom-up and top-down processes in the perception of biological motion.

Second, the mirror neuron system does not account for all aspects of action understanding. As Gallese and Sinigaglia (2014, p. 200) pointed out, action understanding "involves representing to which . . . goals the action is directed; identifying which beliefs, desires, and intentions specify reasons explaining why the action happened; and realising how those reasons are linked to the agent and to her action".

Third, nearly all studies on the mirror neuron system have investigated its properties with respect only to hand actions. However, somewhat differ-ent mirror neuron networks are probably associated with hand-and-mouth actions (Ferrari et al., 2017a).

Fourth, it follows from theoretical approaches to the mirror neuron system that an observer's ability to understand another person's actions should be greater if they both execute any given action in a similar fashion. This prediction has been confirmed (Macerollo et al., 2015). Such research indicates the importance of studying individual differences in motor actions, which have so far been relatively neglected.

CHANGE BLINDNESS

We have seen that a changing visual environment allows us to move in the appropriate direction and to make coherent sense of our surroundings. However, as we will see, our perceptual system does not always respond appropriately to changes in the visual environment.

Have a look around you (go on!). You probably have a strong impres-sion of seeing a vivid and detailed picture of the visual scene. As a result, you are probably confident you could immediately detect any reasonably large change in the visual environment. In fact, that is often *not* the case.

Change blindness
Failure to detect various changes (e.g. in objects) in the visual environment.

Inattentional blindness
Failure to detect an unexpected object appearing in the visual environment.

Change blindness blindness
The tendency of observers to overestimate greatly the extent to which they can detect visual changes and so avoid **change blindness**.

Change blindness, which is "the failure to detect changes in visual scenes" (Ball et al., 2015, p. 2253) is the main phenomenon we will discuss. We also consider **inattentional blindness**, "the failure to consciously perceive otherwise salient events when they are not attended" (Ward & Scholl, 2015, p. 722). Research on change blindness focuses on dynamic processes over time. It has produced striking and counterintuitive findings leading to new theoretical thinking about the processes underlying conscious visual awareness.

Change blindness and inattentional blindness both depend on a mixture of perceptual and attentional processes. It is thus appropriate to discuss these phenomena at the end of our coverage of perception and immediately prior to the start of our coverage of attention.

You have undoubtedly experienced change blindness at the movies caused by unintended continuity mistakes when a scene has been reshot. For example, in the film *Skyfall*, James Bond is followed by a white car. Mysteriously, this car suddenly becomes black and then returns to being white! For more examples, type "Movie continuity mistakes" into YouTube.

We greatly exaggerate our ability to detect visual changes. Levin et al. (2002) asked observers to watch videos involving two people in a restaurant. In one video, the plates change from red to white and in another a scarf worn by one person disappeared. Levin et al. found 46% of observers claimed they would have noticed the change in the colour of the plates without being forewarned and the figure was 78% for the disappearing scarf. In a previous study, 0% of observers detected either change! Levin et al. introduced the term **change blindness blindness** to describe our wildly optimistic beliefs about our ability to detect visual changes.

In the real world, we are often aware of visual changes because we detect motion signals accompanying the change. Laboratory researchers have used various ways to prevent observers from detecting motion signals. One way is to make the change during a saccade (rapid movement of the eyes). Another way is to have a short gap between the original and changed displays (the flicker paradigm).

Suppose you walked across a large square close to a unicycling clown wearing a vivid purple and yellow outfit, large shoes and a bright red nose (see Figure 4.11). Would you spot him? I imagine your answer is "Yes". However, Hyman et al. (2009) found only 51% of people walking on their own spotted the clown. Those failing to spot the clown exhibited inattentional blindness.

Figure 4.11
The unicycling clown who cycled close to students walking across a large square.
From Hyman et al. (2009).

Change blindness vs inattentional blindness

Change blindness and inattentional blindness both involve a failure to detect some visual event

occurring in plain sight. Unsurprisingly, failures of attention often play an important role in causing both forms of blindness.

However, there are major differences between the two phenomena (Jensen et al., 2011). First, consider the effects of instructing observers to look for unexpected objects or visual changes. Target detection in change blindness paradigms is often hard even with such instructions. In contrast, target detection in inattentional blindness paradigms becomes trivially easy. Second, change blindness involves the use of *memory* to compare pre-change and post-change stimuli, whereas inattentional blindness does not. Third, inattentional blindness mostly occurs when the observer's attention is engaged in a demanding task (e.g., chatting on a mobile phone) unlike change blindness.

In sum, more *complex* processing is typically required for successful performance in change blindness tasks. More specifically, observers must engage successfully in *five* separate processes for change detection to occur (Jensen et al., 2011):

(1) Attention must be paid to the change location.
(2) The pre-change visual stimulus at the change location must be encoded into memory.
(3) The post-change visual stimulus at the change location must be encoded into memory.
(4) The pre- and post-change representations must be compared.
(5) The discrepancy between the pre- and post-change representations must be recognised at the conscious level.

IN THE REAL WORLD: IT'S MAGIC! 🌐

Magicians benefit from the phenomena of inattentional blindness and change blindness (Kuhn & Martinez, 2012). Most magic tricks involve *misdirection* which is designed "to disguise the method and thus prevent the audience from detecting it" (Kuhn & Martinez, 2012, p. 2). Many people believe misdirection involves the magician manipulating the audience's attention away from some action crucial to the trick's success. That is often (but not always) the case.

Inattentional blindness

Kuhn and Findlay (2010) studied inattentional blindness using a disappearing lighter (see Figure 4.12 for details). There were three main findings. First, of the observers who detected the drop, 31% were fixating close to the magician's left hand when the lighter was dropped from that hand. However, 69% were fixating some distance away and so detected the drop in peripheral vision (see Figure 4.13). Second, the average distance between fixation and the drop was the same in those who detected the drop in peripheral vision and those who did not. Third, the time taken *after* the drop to fixate the left hand was much less in observers using peripheral vision to detect the drop than those failing to detect it (650 ms vs 1,712 ms).

What do the above findings mean? The lighter drop can be detected by *overt* attention (attention directed to the fixation point) or *covert* attention (attention directed away from the fixation point). Covert attention was surprisingly effective because the human visual system can readily detect movement in peripheral vision (see Chapter 2).
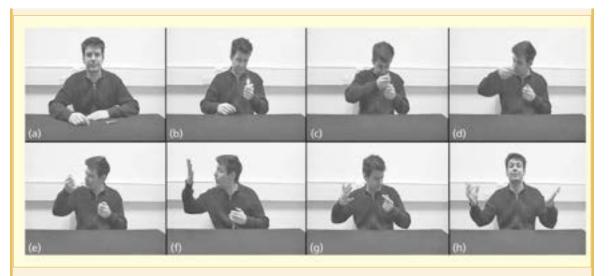
Figure 4.12

The sequence of events in the disappearing lighter trick: (a) the magician picks up a lighter with his left hand and (b) lights it; (c) and (d) he pretends to take the flame with his right hand and (e) gradually moves it away from the hand holding the lighter; (f) he reveals his right hand is empty while the lighter is dropped into his lap; (g) the magician directs his gaze to his left hand and (h) reveals that his left hand is also empty and the lighter has disappeared.

From Kuhn and Findlay (2010). Reprinted with permission of Taylor & Francis.

Most people underestimate the importance of peripheral vision to trick detection. Across several magic tricks (including the lighter trick and other tricks involving change blindness), Ortega et al. (2018) found under 30% of individuals thought they were likely to detect how a trick worked using peripheral vision. In fact, however, over 60% of the tricks where they detected the method involved peripheral vision! Thus, most people exaggerate the role of central vision in understanding magic tricks.

Change blindness

Smith et al. (2012) used a magic trick in which a coin was passed from one hand to the other and then dropped. Observers guessed whether the coin landed heads or tails. On one trial, the coin was switched (e.g., from a £1 coin to a 2p coin). All observers fixated the coin throughout the time it was visible but about 90% failed to detect the coin had changed! Thus, an object can be attended to without some of the features irrelevant to the current task being processed sufficiently to prevent change blindness.

▲ participants who detected the drop
○ participants who missed the drop

Figure 4.13

Participants' fixation points at the time of dropping the lighter for those detecting the drop (triangles) and those missing the drop (circles).

From Kuhn and Findlay (2010). Reprinted with permission of Taylor & Francis.

Kuhn et al. (2016) used a trick in which a magician made the colour of playing cards change. Explicit instructions to observers to keep their eyes on the cards influenced overt attention but failed to reduce change blindness.

Conclusions

The success of many magic tricks depends less on *where* observers are fixating (overt attention) than we might think. Observers can be deceived even when their overt attention is directed to the crucial location. In addition, they often avoid change blindness or inattentional blindness even when their overt attention is directed some distance away from the crucial location. Such findings are typically explained by assuming the focus of covert attention often differs from that of overt attention. More generally, peripheral vision is often of more importance to the detection of magic tricks than most people believe.

Change blindness underestimates visual processing

Ball and Busch (2015) distinguished between two types of change detection: (1) *seeing* the object that changed; (2) *sensing* there has been a change without conscious awareness of which object has changed. Several coloured objects were presented in pre- and post-change displays. If the post-change display contained a colour not present in the pre-change display, observers often sensed change had occurred without being aware of *what* had changed.

When observers show change blindness, it does *not* necessarily mean there was no processing of the change. Ball et al. (2015) used object changes where the two objects were semantically related (e.g., *rail car* changed to *rail*) or unrelated (e.g., *rail car* changed to *sausage*). Use of event-related potentials (ERPs; see Glossary) revealed a larger negative wave when the objects were semantically unrelated even when observers exhibited change blindness. Thus, there was much unconscious processing of the pre- and post-change objects.

What causes change blindness?

There is no single (or simple) answer to the question "What causes change blindness?". However, two major competing theories both provide partial answers. First, there is the attentional approach (e.g., Rensink et al., 1997). According to this approach, change detection requires selective attention to be focused on the object that changes. Attention is typically directed to only a limited part of visual space, and changes in unattended objects are unlikely to be detected.

Second, there is a theoretical approach emphasising the importance of peripheral vision (Rosenholtz, 2017a,b; Sharan et al., 2016 unpublished). It is based on the assumption that visual processing occurs in parallel across the entire visual field (including peripheral vision). According to this approach, "Peripheral vision is a limiting factor underlying standard demonstrations of change blindness" (Sharan et al., 2016, p. 1).

Attentional approach

Change blindness often depends on attentional processes. We typically attend to regions of a visual scene likely to contain salient or important information. Spot the differences between the pictures in Figure 4.14. Observers took an average of 10.4 seconds with the first pair of pictures but only 2.6 with the second pair (Rensink et al., 1997). The height of the railing is less important than the helicopter's position.

Hollingworth and Henderson (2002) recorded eye movements while observers viewed visual scenes (e.g., kitchen; living room). It was assumed the object fixated at any moment was the being attended. There were two potential changes in each scene:

* *Type change*: an object was replaced by one from a different category (e.g., a plate replaced by a bowl).
* *Token change*: an object was replaced by an object from the same category (e.g., a plate replaced by a different plate).

(a) Change in marginal interest (MI)

(b) Change in central interest (CI)
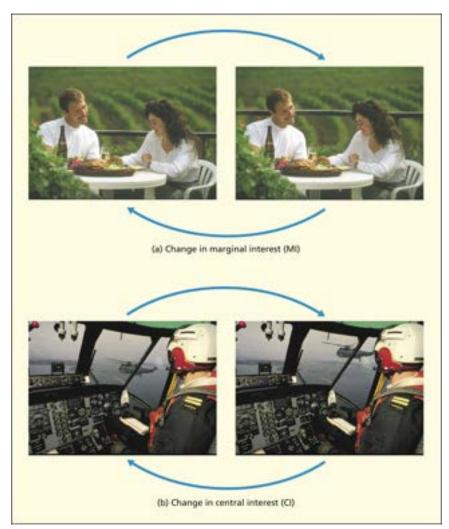
Figure 4.14
(a) The object that is changed (the railing) undergoes a shift in location comparable to that of the object that is changed (the helicopter) in (b). However, the change is much easier to see in (b) because the changed object is more important.

From Rensink et al. (1997). Copyright 1997 by SAGE. Reprinted by permission of SAGE Publications.

What did Hollingworth and Henderson (2002) find? First, there was much greater change detection when the changed object was fixated prior to the change than when it was not fixated (see Figure 4.15a). Second, there was change blindness for 60% of objects fixated prior to changing. Thus, attention to the to-be-changed object was necessary (but not sufficient) for change detection. Third, change detection was much greater when the object type changed rather than simply token change because type changes are more dramatic and obvious.

Evaluation

The attentional approach has various successes to its credit. First, change detection is greater when target stimuli are salient or important and so attract attention. Second, change detection is generally greater when the to-be-changed object has been fixated (attended to) prior to the change.

What are the limitations with the attentional approach? First, the notion that narrow-focused attention determines our visual experience is
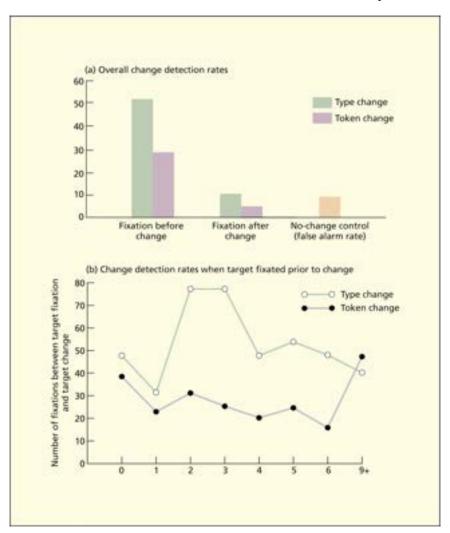
Figure 4.15
(a) Percentage of correct change detection as a function of form of change (type vs token) and time of fixation (before vs after change); also false alarm rate when there was no change. (b) Mean percentage correct change detection as a function of the number of fixations between target fixation and change of target and form of change (type vs token).

Both from Hollingworth and Henderson (2002). Copyright 2002 American Psychological Association. Reproduced with permission.

hard to reconcile with our strong belief that experience spans the entire field of view (Cohen et al., 2016). Second, "A selective attention account is hard to prove or disprove, as it relies on largely unknown attentional loci as well as poorly understood effects of attention" (Sharan et al., 2016).

Third, change blindness is sometimes poorly predicted by the focus of overt attention (indexed by eye fixations) (e.g., Smith et al., 2012; Kuhn et al., 2016). Such findings are often explained by covert attention, but this is typically not measured directly. Fourth, the attentional approach implies incorrectly that very little useful information is extracted from visual areas outside the focus of attention (see below).

Peripheral vision approach

Visual acuity is greatest in the centre of the visual field (the fovea; see Glossary). However, peripheral vision (all vision outside the fovea) typically covers the great majority of the visual field (see Chapter 2). As Rosenholtz (2016, p. 438) pointed out, it is often assumed "Peripheral vision is impoverished and all but useless". This is a great exaggeration even though acuity and colour perception are much worse in the periphery than the fovea. In fact, peripheral vision is often most impaired by **visual crowding**: "identification of a peripheral object is impaired by nearby objects" (Pirkner & Kimchi, 2017, p. 1) (see Chapter 5).

According to Sharan et al. (2016, p. 3), "The hypothesis that change blindness may arise in part from limitations of peripheral vision is quite different from usual explanations of the phenomenon [which attribute it to] a mix of inattention and lack of details stored in memory."

Sharan et al. (2016) tested the above hypothesis. Initially, they categorised change-detection tasks as easy, medium and hard on the basis of how rapidly observers detected the change. Then they presented these tasks to different observers who fixated at various degrees of visual angle (eccentricities) from the area that changed. There were two key findings:

(1) Change-detection performance was surprisingly good even when the change occurred well into peripheral vision.
(2) Peripheral vision plays a major role in determining change-detection performance – hard-to-detect changes require closer fixations than those that are easy to detect.
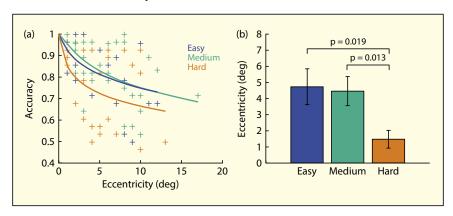
Figure 4.16
(a) Change-detection accuracy as a function of task difficulty and visual eccentricity. (b) The eccentricity at which change-detection accuracy was 85% correct as a function of task difficulty.

From Sharan et al. (2016).

Further evidence that much information is extracted from peripheral as well as foveal vision was reported by Clarke and Mack (2015). On each trial, two real-world scenes were presented with an interval of 1,500 ms between them. When this interval was unfilled, only 11% of changes were detected. However, when a cue indicating the location of a possible change was presented 0, 300 or 1,000 ms after the offset of the first scene, change-detection rates were much higher. They were greatest in the 0 ms condition (29%) and lowest in the 1,000 ms condition (18%). Thus, much information about the first scene (including information from peripheral vision) was stored briefly in iconic memory (see Glossary and Chapter 6).

If peripheral vision provides observers with general or gist information, they might detect global scene changes without detecting precisely *what* had changed. Howe and Webb (2014) obtained support for this prediction. Observers were presented with an array of 30 discs (15 red, 15 green). On 24% of trials when three discs changed colour, observers detected the array had changed but could not identify the discs involved.

Evaluation

The peripheral vision approach has proved successful in various ways. First, visual information is often extracted from across the entire visual field as predicted by this approach (but not the attentional approach). This supports our strong belief that we perceive most of the immediate visual environment. Second, this approach capitalises on established knowledge concerning peripheral vision. Third, this approach has been applied successfully to explain visual-search performance (see Chapter 5).

What are the limitations of this approach? First, it de-emphasises attention's role in determining change blindness, and does not provide a detailed account of how attentional and perceptual processes are integrated. Second, Sharan et al. (2016) discovered change detection was sometimes difficult even though the change could be perceived easily in peripheral vision. This indicates that other factors (as yet unidentified) are also involved. Third, the approach does not consider failure to compare pre- and post-change representations as a reason for change blindness (see below).

Comparison of pre- and post-change representations

Change blindness can occur because observers fail to *compare* their pre- and post-change representations. Angelone et al. (2003) presented a video in which the identity of the central actor changed. On a subsequent line-up task to identify the pre-change actor, observers showing change blindness performed comparably to those showing change detection (53% vs 46%, respectively).

Varakin et al. (2007) extended the above research in a real-world study in which a coloured binder was switched for one of a different colour while observers' eyes were closed. Some observers exhibited change blindness even though they remembered the colours of the pre- and post-change binders and so had failed to compare the two colours. Other observers showing change blindness had poor memory for the pre- and post-change colours and so failed to represent these two pieces of information in memory.

KEY TERM 🔑

Serial dependence
Systematic bias of current visual perception towards recent visual input.

Is change blindness a defect?

Is change blindness an unfortunate defect? Fischer and Whitney (2014) argued the answer is "No". The visual world is typically relatively stable over short time periods. As a result, it is worthwhile for us to sacrifice perceptual accuracy occasionally to ensure we have a continuous, stable perception of our visual environment.

Fischer and Whitney (2014) supported their argument by finding the perceived orientation of a grating was biased in the direction of a previously presented grating, an effect known as serial dependence. Manassi et al. (2018) found serial dependence for an object's location – when an object that had been presented previously was re-presented, it was perceived as being closer to its original location than was actually the case. Serial dependence probably involves several stages of visual perception and may also involve memory processes (Bliss et al., 2017). In sum, the visual system's emphasis on perceptual stability inhibits our ability to detect changes within the visual scene.

Inattentional blindness and its causes

The most famous study on inattentional blindness was reported by Simons and Chabris (1999). In one condition, observers watched a video where students dressed in white (the white team) passed a ball to each other and the observers counted the number of passes (see the video at www.simonslab.com/videos.html). At some point, a woman in a black gorilla suit walks into camera shot, looks at the camera, thumps her chest and then walks off (see Figure 4.17). Altogether she is on screen for 9 seconds. Very surprisingly, only 42% of observers noticed the gorilla! This is a striking example of inattentional blindness.

Figure 4.17
Frame showing a woman in a gorilla suit in the middle of a game of passing the ball.

From Simons & Chabris (1999). Figure provided by Daniel Simons, www.dansimons.com/www.theinvisiblegorilla.com.

Why was performance so poor in the above experiment? Simons and Chabris (1999) obtained additional relevant evidence. In a second condition, observers counted the number of passes made by students dressed in black. Here 83% of observers detected the gorilla's presence. Thus, observers were more likely to attend to the gorilla when it resembled task-relevant stimuli (i.e., in colour).

It is generally assumed detection performance is good when observers count black team passes because of selective attention to black objects. Indeed, Rosenholtz et al. (2016) found that observers counting black team passes had eye fixations closer to the gorilla than those counting white team passes. However, Rosenholtz et al. also found that observers counting black team passes (but whose fixation patterns resembled those of observers counting white

team passes) had unusually poor detection performance (54% compared to a typical 80%). Thus, detection performance may depend on the strengths and limitations of peripheral vision as well as failures of selective attention.

The presence of inattentional blindness can lead us to *underestimate* the amount of processing of the undetected stimulus. Schnuerch et al. (2016) found categorising attended stimuli was slower when the meaning of an undetected stimulus conflicted with that of the attended stimulus. Thus, the meaning of undetected stimuli was processed despite inattentional blindness. Other research using event-related potentials (reviewed by Pitts, 2018) has shown that undetected stimuli typically receive moderate processing.

How can we explain inattentional blindness? As we have seen, explanations often emphasise the role of selective attention or attentional set. Simons and Chabris' (1999) findings indicate the importance of similarity in stimulus *features* (e.g., colour) between task stimuli and the unexpected object. However, Most (2013) argued that similarity in *semantic category* is also important. Participants tracked numbers or letters. On the critical trial, an unexpected stimulus (the letter *E* or number *3*) was visible for 7 seconds. The letter and number were visually identical except they were mirror images of each other.

What did Most (2013) find? There was much less inattentional blindness when the unexpected stimulus belonged to the same category as the tracked objects. Thus, inattentional blindness can depend on attentional sets based on semantic categories (e.g., letters; numbers).

Légal et al. (2017) investigated the role of demanding top-down attentional processes in producing inattentional blindness using Simons and Chabris' (1999) gorilla video. Some observers counted the passes made by the white team (standard task) whereas others had the more attentionally demanding task of counting the number of aerial passes as well as total passes. As predicted, there was much more evidence of inattentional blindness (i.e., failing to detect the gorilla) when the task was more demanding.
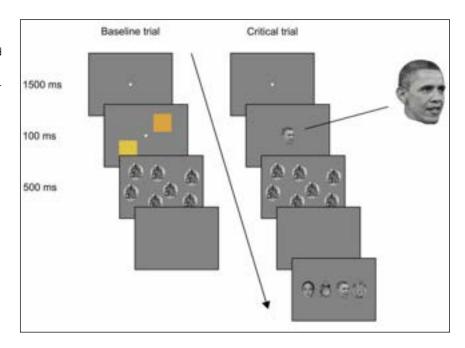
Légal et al. (2017) reduced inattentional blindness in other conditions by presenting detection-relevant words subliminally (e.g., *identify*; *notice*) to observers *prior* to watching the video. This increased detection rates for the gorilla in the standard task condition from 50% to 83%. Overall, the findings indicate that manipulating attentional processes can have powerful effects on inattentional blindness.

Compelling evidence that inattentional blindness depends on top-down processes that strongly influence what we expect to see was reported by Persuh and Melara (2016). Observers fixated a central dot followed by the presentation of two coloured squares and decided whether the colours were the same. On the critical trial, the dot was replaced by Barack Obama's face (see Figure 4.18). Amazingly, 60% of observers failed to detect this unexpected stimulus presented in foveal vision: Barack Obama blindness. Of these observers, a below-chance 8% identified Barack Obama when deciding whether the unexpected stimulus was Angelina Jolie, a lion's head, an alarm clock or Barack Obama (see Figure 4.18).

Persuh and Melara's (2016) findings are dramatic because they indicate inattentional blindness can occur even when the novel stimulus is presented on its own with no competing stimuli. These findings suggest there

Figure 4.18
The sequence of events on the initial baseline trials and the critical trial.

From Persuh and Melara (2016).

are important differences in the processes underlying inattentional blindness and change blindness: the latter often depends on visual crowding (see Glossary), which is totally absent in Persuh and Melara's study.

Evaluation

Several factors influencing inattentional blindness have been identified. These factors include the similarity (in terms of stimulus features and semantic category) between task stimuli and the unexpected object; the attentional demands of the task; and observers' expectations concerning what they will see. If there were no task requiring attentional resources and creating expectations, there would undoubtedly be very little inattentional blindness (Jensen et al., 2011).

What are the limitations of research in this area? First, it is typically unclear whether inattentional blindness is due to perceptual failure or to memory failure (i.e., the unexpected object is perceived but rapidly forgotten). However, Ward and Scholl (2015) found that observers showed inattentional blindness even when observers were instructed to report immediately seeing anything unexpected. This finding strongly suggests that inattentional blindness reflects deficient perception rather than memory failure.

Second, observers typically engage in some processing of undetected stimuli even when they fail to report the presence of such stimuli (Pitts, 2018). More research is required to clarify the extent of non-conscious processing of undetected stimuli.

Third, it is likely that the various factors influencing inattentional blindness interact in complex ways. However, most research has considered only a single factor and so the nature of such interactions has not been established.

CHAPTER SUMMARY 📝

- **Introduction.** The time dimension is very important in visual perception. The changes in visual perception produced as we move around the environment and/or environmental objects move promote accurate perception and facilitate appropriate actions.

- **Direct perception.** Gibson argued perception and action are closely intertwined and so research should not focus exclusively on static observers perceiving static visual displays. According to his direct theory, an observer's movement creates optic flow providing useful information about the direction of heading. Invariants, which are unchanged as people move around their environment, have particular importance. Gibson claimed the uses of objects (their affordances) are perceived directly. He underestimated the complexity of visual processing, minimising the role of object knowledge in visual perception, with the effects of motion on perception being more complex than he realised.

- **Visually guided movement.** The perception of heading depends in part on optic-flow information. However, there are complexities because the retinal flow field is determined by eye and head movements as well as by optic flow. Heading judgements are also influenced by binocular disparity and the retinal displacement of objects as we approach them.

  Accurate steering on curved paths (e.g., driving around a bend) sometimes involves focusing on the tangent point (e.g., point on the inside edge of the road at which its direction seems to reverse). However, drivers sometimes fixate a point along the future path. More generally, drivers' gaze patterns are flexibly determined by control mechanisms that are responsive to their goals.

  Calculating time to contact with an object often involves calculating tau (the size of the retinal image divided by the object's rate of expansion). Drivers often use tau-dot (rate of decline of tau over time) to decide whether there is sufficient braking time to stop before contact. Observers often make use of additional sources of information (e.g., binocular disparity; familiar size; relative size) when working out time to contact. Drivers' braking decisions also depend on their preferred margin of safety and the effectiveness of the car's braking system.

- **Visually guided action: contemporary approaches.** The planning-control model distinguishes between a slow planning system used mostly before the initiation of movement and a fast control system used during movement execution. As predicted, separate brain areas are involved in planning and control. However, the definition of "planning" is very broad, and the notion that planning always precedes control is oversimplified. Recent evidence indicates

that visually guided action depends on three processing streams (dorso-dorsal; the ventro-dorsal; and ventral, which is discussed more fully in Chapter 2) each making a separate contribution. This theoretical approach is supported by studies on brain-damaged patients and by neuroimaging research.

•   **Perception of human motion.** Human motion is perceived even when only impoverished visual information is available. Perception of human and biological motion involves bottom-up and top-down processes with the latter most likely to be used with degraded visual input. The perception of human motion is special because we can produce as well as perceive human actions and because we devote considerable time to making sense of it.

It has often been assumed that our ability to imitate and understand human motion depends on a mirror neuron system (an extensive brain network). This system's causal involvement in action perception and understanding has been shown in research on brain-damaged patients and studies using techniques to alter its neural activity. The mirror neuron system is especially important in the understanding of relatively simple actions. However, additional high-level cognitive processes are often required if action understanding is complex or involves generalising from past experience.

•   **Change blindness.** There is convincing evidence for change blindness and inattentional blindness. Change blindness depends on attentional processes: it occurs more often when the changed object does not receive attention. However, change blindness can occur for objects that are fixated and it also depends on the limitations of peripheral vision. The visual system's emphasis on continuous, stable perception probably plays a part in making us susceptible to change blindness. Inattentional blindness depends very strongly on top-down processes (e.g., selective attention) and can be found even when only the novel stimulus is present in the visual field.

FURTHER READING

Binder, E., Dovern, A., Hesse, M.D., Ebke, M., Karbe, H., Salinger, J. et al. (2017). Lesion evidence for a human mirror neuron system. *Cortex*, *90*, 125–137. Ellen Binder and colleagues discuss the nature of the mirror neuron system based on evidence from brain-damaged patients.

Keysers, C., Paracampo, R. & Gazzola, V. (2018). What neuromodulation and lesion studies tell us about the function of the mirror neuron system and embodied cognition. *Current Opinion in Psychology*, *24*, 35–40. This article provides a succinct account of our current understanding of the mirror neuron system.

Lappi, O. & Mole, C. (2018). Visuo-motor control, eye movements, and steering: A unified approach for incorporating feedback, feedforward, and internal models. *Psychological Bulletin*, *144*, 981–1001. Otto Lappi and Callum Mole provide

a comprehensive theoretical account of driving behaviour that emphasises the importance of top-down control mechanisms in influencing drivers' eye fixations.

Osiurak, F., Rossetti, Y. & Badets, A. (2017). What is an affordance? 40 years later. *Neuroscience and Biobehavioral Reviews*, *77*, 403–417. François Osiurak and colleagues discuss Gibson's notion of affordances in the contest of contemporary research and theory.

Rosenholtz, R. (2017a). What modern vision science reveals about the awareness puzzle: Summary-statistic encoding plus decision limits underlie the richness of visual perception and its quirky failures. *Vision Sciences Society Symposium on Summary Statistics and Awareness*, preprint arXiv:1706.02764. Ruth Rosenholtz provides an excellent account of the role played by peripheral vision in change blindness and other phenomena.

Sakreida, K., Effnert, I., Thill, S., Menz, M.M., Jirak, D., Eickhoff, C.R. et al. (2016). Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience and Biobehavioral Reviews*, *69*, 80–112. The pathways within the brain involved in goal-directed interactions with objects are discussed in the context of a meta-analytic review.

Chapter
5

Attention and performance

INTRODUCTION

Attention is invaluable in everyday life. We use attention to avoid being hit by cars when crossing the road, to search for missing objects and to perform two tasks together. The word "attention" has various meanings but typically refers to *selectivity* of processing as emphasised by William James (1890, pp. 403–404):

> Attention is . . . the taking into possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation, concentration, of consciousness are of its essence.

William James distinguished between "active" and "passive" modes of attention. Attention is active when controlled in a top-down way by the individual's goals or expectations. In contrast, attention is passive when controlled in a bottom-up way by external stimuli (e.g., a loud noise). This distinction remains theoretically important (e.g., Corbetta & Shulman, 2002; see discussion, pp. 192–196).

Another important distinction is between focused and divided attention. **Focused attention** (or selective attention) is studied by presenting individuals with two or more stimulus inputs at the same time and instructing them to respond to only one. Research on focused or selective attention tells us how effectively we can *select* certain inputs and avoid being distracted by non-task inputs. It also allows us to study the selection process and the fate of unattended stimuli.

Divided attention is also studied by presenting at least two stimulus inputs at the same time. However, individuals are instructed they must attend (and respond) to *all* stimulus inputs. Divided attention is also known as multi-tasking (see Glossary). Studies of divided attention provide useful information about our processing limitations and the capacity of our attentional mechanisms.

There is a final important distinction (the last one, I promise you!) between *external* and *internal* attention. External attention is "the selection and modulation of sensory information" (Chun et al., 2011). In contrast, internal attention is "the selection, modulation, and maintenance of internally generated information, such as task rules, responses, long-term memory, or working memory" (Chun et al., 2011, p. 73). The connection to Baddeley's working memory model is especially important (e.g., Baddeley (2012); see Chapter 6). The central executive component of working memory is involved in attentional control and is crucially involved in internal and external attention.

Much attentional research has two limitations. First, the emphasis is on attention to externally presented task stimuli rather than internally generated stimuli (e.g., worries; self-reflection). One reason is that it is easier to assess and to control external attention. Second, what participants attend to is determined by the experimenter's instructions. In contrast, what we attend to in the real world is mostly determined by our current goals and emotional states.

Two important topics related to attention are discussed elsewhere. Change blindness (see Glossary), which shows the close links between attention and perception, is considered in Chapter 4. Consciousness (including its relationship to attention) is discussed in Chapter 16.

FOCUSED AUDITORY ATTENTION

Many years ago, British scientist Colin Cherry (1953) became fascinated by the **cocktail party problem** – how can we follow just one conversation when several people are talking at once? As we will see, there is no simple answer.

McDermott (2009) identified two problems listeners face when attending to one voice among many. First, there is *sound segregation*: the listener must decide *which* sounds belong together. This is complex: machine-based speech recognition programs often perform poorly when attempting to achieve sound segregation with several sound sources present together (Shen et al., 2008). Second, after segregation has been achieved, the listener must direct attention to the *sound source of interest* and ignore the others.

McDermott (2009) pointed out that auditory segmentation is often harder than visual segmentation (deciding which visual features belong to which objects; see Chapter 3). There is considerable overlap of signals from different sound sources in the cochlea whereas visual objects typically occupy different retinal regions.

There is another important issue – when listeners attend to one auditory input, how much processing is there of the unattended input(s)? As we will see, various answers have been proposed.

Cherry (1953) addressed the issues discussed so far (see Eysenck, 2015, for an evaluation of his research). He studied the cocktail party problem using a **dichotic listening task** in which a different auditory message was presented to each ear and the listener attended to only one. Listeners engaged in **shadowing** (repeating the attended message aloud as it was presented) to ensure their attention was directed to that message. However, the shadowing task has two potential disadvantages: (1) listeners do not

normally engage in shadowing and so the task is artificial; and (2) it increases listeners' processing demands.

Listeners solved the cocktail party problem by using *differences* between the auditory inputs in physical features (e.g., sex of speaker; voice intensity; speaker location). When these physical differences were eliminated by presenting two messages in the same voice to both ears at once, listeners found it very hard to separate out the messages based on differences in meaning.

Cherry (1953) found very little information seemed to be extracted from the unattended message. Listeners seldom noticed when it was spoken backwards or in a foreign language. However, physical changes (e.g., a pure tone) were nearly always detected. The conclusion that unattended information receives minimal processing was supported by Moray (1959), who found listeners remembered very few words presented 35 times each.

Where is the bottleneck? Early vs late selection

Many psychologists have argued we have a processing bottleneck (discussed below). A bottleneck in the road (e.g., where it is especially narrow) can cause traffic congestion, and a bottleneck in the processing system seriously limits our ability to process two (or more) simultaneous inputs. However, it would sometimes solve the cocktail problem by permitting listeners to process only the desired voice.

Where is the bottleneck? Broadbent (1958) argued a filter (bottleneck) early in processing allows information from one input or message through it based on the message's physical characteristics. The other input remains briefly in a sensory buffer and is rejected unless attended to rapidly (see Figure 5.1). Thus, Broadbent argued there is *early* selection.

Treisman (1964) argued the bottleneck's location is more *flexible* than Broadbent suggested (see Figure 5.1). She claimed listeners start with processing based on physical cues, syllable pattern and specific words and then process grammatical structure and meaning. Later processes are omitted or attenuated if there is insufficient processing capacity to permit full stimulus analysis.
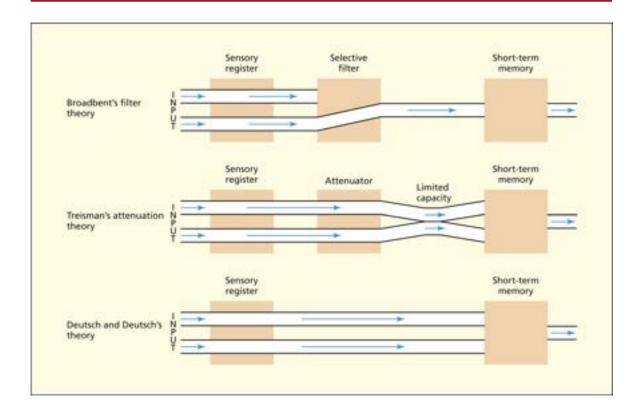
Treisman (1964) also argued top-down processes (e.g., expectations) are important. Listeners performing the shadowing task sometimes say a word from the unattended input. Such breakthroughs mostly occur when the word on the unattended channel is highly probable in the context of the attended message.

Deutsch and Deutsch (1963) argued all stimuli are fully analysed, with the most important or relevant stimulus determining the response. Thus, they placed the bottleneck much later in processing than did Broadbent (see Figure 5.1).

Findings: unattended input

Broadbent's approach predicts little or no processing of unattended auditory messages. In contrast, Treisman's approach suggests flexibility in the processing of unattended messages, whereas Deutsch and Deutsch's approach implies reasonably thorough processing of such messages. Relevant findings are discussed below.

Interactive exercise:
Treisman

Treisman and Riley (1969) asked listeners to shadow one of two auditory messages. They stopped shadowing and tapped when they detected a target in either message. Many more target words were detected on the shadowed message.

Aydelott et al. (2015) asked listeners to perform a task on attended target words. When unattended words related in meaning were presented shortly before the target words themselves, performance on the target words was enhanced when unattended words were presented as loudly as attended ones. Thus, the meaning of unattended words was processed.

There is often more processing of unattended words that have a special significance for the listener. For example, Li et al. (2011) obtained evidence that unattended weight-related words (e.g., *fat*; *chunky*) were processed more thoroughly by women dissatisfied with their weight. Conway et al. (2001) found listeners often detected their own name on the unattended message. This was especially the case if they had low working memory capacity (see Glossary) indicative of poor attentional control.

Coch et al. (2005) asked listeners to attend to one of two auditory inputs and to detect targets presented on either input. Event-related potentials (ERPs; see Glossary) provided a measure of processing activity. ERPs 100 ms after target presentation were greater when the target was presented on the attended rather than the unattended message. This suggests there was more processing of the attended than unattended targets.

Greater brain activation for attended than unattended auditory stimuli may reflect *enhanced* processing for attended stimuli and/or *suppressed* processing for unattended stimuli. Horton et al. (2013) addressed this

Figure 5.1
A comparison of Broadbent's theory (top), Treisman's theory (middle), and Deutsch and Deutsch's theory (bottom).

issue. Listeners heard separate speech messages presented to each ear with instructions to attend to the left or right ear. There was greater brain activation associated with the attended message (especially around 90 ms after stimulus presentation). This difference depended on *enhancement* of the attended message combined with *suppression* of the unattended message.

Classic theories of selective auditory attention (those of Broadbent, Treisman, and Deutsch and Deutsch) de-emphasised the importance of suppression or inhibition of the unattended message shown by Horton et al. (2013). For example, Schwartz and David (2018) reported suppression of neuronal responses in the primary auditory cortex to distractor sounds. More generally, all the classic theories de-emphasise the *flexibility* of selective auditory attention and the role of top-down processes in selection (see below).

Findings: cocktail party problem

Humans are generally very good at separating out and understanding one voice from several speaking at the same time (i.e., solving the cocktail party problem). The extent of this achievement is indicated by the finding that automatic speech recognition systems are considerably inferior to human speech recognition (Spille & Meyer, 2014).

Mesgarani and Chang (2012) studied listeners with implanted multi-electrode arrays permitting the direct recording of activity within the auditory cortex. They heard two different messages (one in a male voice; one in a female voice) presented to the same ear with instructions to attend to only one. The responses within the auditory cortex revealed "The salient spectral [based on sound frequencies] and temporal features of the attended speaker, as if subjects were listening to that speaker alone" (Mesgarani & Chang, 2012, p. 233).

Listeners found it easy to distinguish between the two messages in the study by Mesgarani and Chang (2012) because they differed in physical characteristics (i.e., male vs female voice). Olguin et al. (2018) presented native English speakers with two messages in different female voices. The attended message was always in English whereas the unattended message was in English or an unknown language. Comprehension of the attended message was comparable in both conditions. However, there was stronger neural encoding of both messages in the former condition. As Olguin et al. concluded, "The results offer strong support to flexible accounts of selective [auditory] attention" (p. 1618).

In everyday life, we are often confronted by several different speech streams. Accordingly, Puvvada and Simon (2017) presented *three* speech streams and assessed brain activity as listeners attended to only one. Early in processing, "the auditory cortex maintains an acoustic representation of the auditory scene with no significant preference to attended over ignored sources" (p. 9195). Later in processing, "Higher-order auditory cortical areas represent an attended speech stream separately from, and with significantly higher fidelity [accuracy] than, unattended speech streams" (p. 9189). This latter finding results from top-down processes (e.g., attention).

How do we solve the cocktail party problem? The importance of top-down processes is suggested by the existence of extensive *descending*

pathways from the auditory cortex to brain areas involved in early auditory processing (Robinson & McAlpine, 2009). Various top-down factors based on listeners' knowledge and/or expectations are involved. For example, listeners are more accurate at identifying what one speaker is saying in the context of several other voices if they have previously heard that speaker's voice in isolation (McDermott, 2009).

Woods and McDermott (2018) investigated top-down processes in selective auditory attention in more detail. They argued, "Sounds produced by a given source often exhibit consistencies in structure that might be useful in separating sources" (p. E3313). They used the term "schemas" to refer to such structural consistencies.

Listeners showed clear evidence of schema learning leading to rapid improvements in their listening performance. An important aspect of such learning is temporal coherence – a given source's sound features are typically all present when it is active and absent when it is silent. Shamma et al. (2011) discussed research showing that if listeners can identify *one* distinctive feature of the target voice, they can then distinguish its other sound features via temporal coherence.

Evans et al. (2016) compared patterns of brain activity when attended speech was presented on its own or together with competing unattended speech. Brain areas associated with attentional and control processes (e.g., frontal and parietal regions) were more activated in the latter condition. Thus, top-down processes relating to attention and control are important in selective auditory processing.

Finally, Golumbic et al. (2013) suggested individuals at actual cocktail parties can potentially use *visual* information to assist them in understanding what a given speaker is saying. Listeners heard two simultaneous messages (one in a male voice and the other in a female voice). Processing of the attended message was enhanced when they saw a video of the speaker talking.

In sum, listeners generally achieve the complex task of selecting one speech message from among several such messages. There has been progress in identifying the top-down processes involved. For example, if listeners can identify at least one consistently distinctive feature of the target voice, this makes it easier for them to attend only to that voice. Top-down processes often produce a "winner-takes-all" situation where the processing of one auditory input (the winner) suppresses the brain activity associated with all other inputs (Kurt et al., 2008).

FOCUSED VISUAL ATTENTION

There has been much more research on visual attention than auditory attention. The main reason is that vision is our most important sense modality with more of the cortex devoted to it than any other sense. Here we consider *four* key issues. First, what is focused visual attention like? Second, what is selected in focused visual attention? Third, what happens to unattended visual stimuli? Fourth, what are the major systems involved in visual attention? In the next section (see pp. 196–200), we discuss what the study of visual disorders has taught us about visual attention.

<div style="float: left; width: 30%">

> **KEY TERM** 🔑
>
> **Split attention**
> Allocation of attention to two (or more) non-adjacent regions of visual space.

</div>

Spotlight, zoom lens or multiple spotlights?

Look around you and attend to any interesting objects. Was your visual attention like a spotlight? A spotlight illuminates a fairly small area, little can be seen outside its beam and it can be redirected to focus on any given object. Posner (1980) argued the same is true of visual attention.

Other psychologists (e.g., Eriksen & St. James, 1986) claim visual attention is more *flexible* than suggested by the spotlight analogy and argue visual attention resembles a zoom lens. We can increase or decrease the area of focal attention just as a zoom lens can be adjusted to alter the visual area it covers. This makes sense. For example, car drivers often need to narrow their attention after spotting a potential hazard.

A third theoretical approach is even more flexible. According to the multiple spotlights theory (Awh & Pashler, 2000), we sometimes exhibit **split attention** (attention directed to two or more non-adjacent regions in space). The notion of split attention is controversial. Jans et al. (2010) argued attention is often strongly linked to motor action and so attending to two separate objects might disrupt effective action. However, there is no strong evidence for such disruption.

Findings

Support for the zoom-lens model was reported by Müller et al. (2003). On each trial, observers saw four squares in a semi-circle and were cued to attend to one, two or all four. Four objects were then presented (one in each square) and observers decided whether a target (e.g., a white circle) was among them. Brain activation in early visual areas was most widespread when the attended region was large (i.e., attend to all four squares) and was most limited when it was small (i.e., attend to one square). As predicted by the zoom-lens theory, performance (reaction times and errors) was best with the smallest attended region and worst with the largest one.

Chen and Cave (2016, p. 1822) argued the optimal attentional zoom setting "includes all possible target locations and excludes possible distractor locations". Most findings indicated people's attentional zoom setting is close to optimal. However, Collegio et al. (2019) obtained contrary findings. Drawings of large objects (e.g., *jukebox*) and small objects (e.g., *watch*) were presented so their retinal size was the same. The observer's area of focal attention was greater with large objects because they made top-down inferences concerning their real-world sizes. As a result, the area of focal attention was larger than optimal for large objects.

Goodhew et al. (2016) pointed out that nearly all research has focused only on spatial perception (e.g., identification of a specific object). They focused on temporal perception (was a disc presented continuously or were there two presentations separated by a brief interval?). Spotlight size had *no* effect on temporal acuity, which is inconsistent with the theory. How can we explain these findings? Spatial resolution is poor in peripheral vision but temporal resolution is good. As a consequence, a small attentional spotlight is more beneficial for spatial than temporal acuity.

We turn now to split attention. Suppose you had to identify two digits that would probably be presented to two cued locations a little way apart
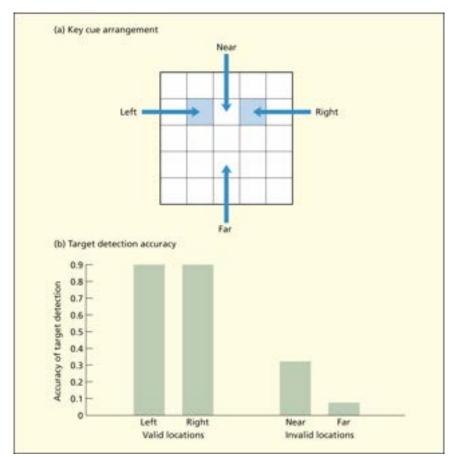
(see Figure 5.2a). Suppose also that on some trials a digit was presented between the two cued locations. According to zoom-lens theory, the area of maximal attention should include the two cued locations *and* the space in between. As a result, the detection of digits presented in the middle should have been very good. In fact, Awh and Pashler (2000) found it was poor (see Figure 5.2b). Thus, attention can resemble multiple spotlights, as predicted by the split-attention approach.

Morawetz et al. (2007) presented letters and digits at five locations simultaneously (one in each quadrant of the visual field and one in the centre). In one condition, observers attended to the visual stimuli at the upper left and bottom right locations and ignored the other stimuli. There were two peaks of brain activation corresponding to the attended areas but less activation corresponding to the region in between. Overall, the pattern of activation strongly suggested split attention.

Niebergall et al. (2011) recorded the neuronal responses of monkeys attending to two moving stimuli while ignoring a distractor. In the key condition, there was a distractor between (and close to) the two attended stimuli. In this condition, neuronal responses to the distractor *decreased* compared to other conditions. Thus, split attention involves a mechanism reducing attention to (and processing of) distractors located between attended stimuli.

Hemifield
One half of the visual field. Information from the left hemifield of each eye proceeds to the right hemisphere and information from the right hemifield proceeds to the left hemisphere.

In most research demonstrating split attention, the two non-adjacent stimuli being attended simultaneously were each presented to a different hemifield (one half of the visual field). Note that the right hemisphere receives visual signals from the left hemifield and the left hemisphere receives signals from the right hemifield. Walter et al. (2016) found performance was better when non-adjacent stimuli were presented to *different* hemifields rather than the *same* hemifield. Of most importance, the assessment of brain activity indicated effective filtering or inhibition of stimuli presented between the two attended stimuli only when presented to different hemifields.

In sum, we can use visual attention very flexibly. Visual selective attention can resemble a spotlight, a zoom lens or multiple spotlights, depending on the current situation and the observer's goals. However, split attention may require that two stimuli are presented to different hemifields rather than the same one. A limitation with all these theories is that metaphors (e.g., attention is a zoom lens) are used to *describe* experimental findings but these metaphors fail to specify the underlying mechanisms (Di Lollo, 2018).

What is selected?

Why might selective attention resemble a spotlight or zoom lens? Perhaps we selectively attend to an area or region of space: *space-based attention*. Alternatively, we may attend to a given object or objects: *object-based attention*. Object-based attention is prevalent in everyday life because visual attention is mainly concerned with objects of interest to us (see Chapters 2 and 3). As expected, observers' eye movements as they view natural scenes are directed almost exclusively to objects (Henderson & Hollingworth, 1999). However, even though we typically focus on objects of potential importance, our attentional system is so *flexible* we can attend to an area of space *or* a given object.

There is also *feature-based attention*. For example, suppose you are looking for a friend in a crowd. Since she nearly always wears red clothes, you might attend to the feature of colour rather than specific objects or locations. Leonard et al. (2015) asked observers to identify a red letter within a series of rapidly presented letters. Performance was impaired when a # symbol also coloured red was presented very shortly before the target. Thus, there was evidence for feature-based attention (e.g., colour; motion).

Findings

Visual attention is often object-based. For example, O'Craven et al. (1999) presented observers with two stimuli (a face and a house), transparently overlapping at the same location, with instructions to attend to one of them. Brain areas associated with face processing were more activated when the face was attended to than when the house was. Similarly, brain areas associated with house processing were activated when the house was the focus of attention.

Egly et al. (1994) devised a much-used method for comparing object-based and space-based attention (see Figure 5.3). The task was to select a target stimulus as rapidly as possible. A cue presented before the target

was valid (same location as the target) or invalid (different location from the target). Of key importance, invalid cues were in the *same* object as the target (within-object cues) or in a *different* object (between-object cues). The key finding was that target detection was faster on invalid trials when the cue was in the same object rather than a different one. Thus, attention was at least partly object-based.

Does object-based attention in the Egly et al. (1994) task occur fairly "automatically" or does it involve strategic processes? Object-based attention should *always* be found if it is automatic. Drummond and Shomstein (2010) found no evidence for object-based attention when the cue indicated with 100% certainty where the target would appear. Thus, any preference for object-based attention can be *overridden* when appropriate.
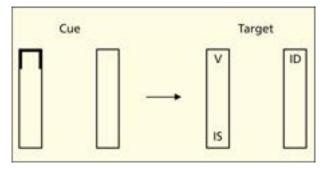
Figure 5.3

Stimuli adapted from Egly et al. (1994). Participants saw two rectangles and a cue indicated the most likely location of a subsequent target. The target appeared at the cued location (V), at the uncued end of the cued rectangle (IS) or at the uncued, equidistant end of the uncued rectangle (ID).

From Chen (2012). © Psychonomic Society, Inc. Reprinted with permission from Springer.

Hollingworth et al. (2012) found evidence object-based *and* space-based attention can occur at the same time using a task resembling that of Egly et al. (1994). There were three types of within-object cues varying in the distance between the cue and subsequent target (see Figure 5.4). There was evidence for object-based attention: when the target was far from the cue, performance was worse when the cue was in a different object rather than the same one. There was also evidence for space-based attention: when the target was in the same object as the cue, performance declined the greater the distance between target and cue. Thus, object-based and space-based attention are not mutually exclusive.

Similar findings were reported by Kimchi et al. (2016). Observers responded faster to a target presented *within* rather than *outside* an object. This indicates object-based attention. There was also evidence for space-based attention: when targets were presented outside the object, observers responded faster when they were close to it. Kimchi et al. concluded that "object-related and space-related attentional processing can operate simultaneously" (p. 48).

Pilz et al. (2012) compared object-based and space-based attention using various tasks. Overall, there was much more evidence of space-based than object-based attention, with only a small fraction of participants showing clear-cut evidence of object-based attention.

Donovan et al. (2017) noted that most studies indicating visual attention is object-based have used spatial cues, which may bias the allocation of attention. Donovan et al. avoided the use of spatial cues and found "Object-based representations do not guide attentional selection in the absence of spatial cues" (p. 762). This finding suggests previous research has exaggerated the extent of object-based visual attention.

When we search the visual environment, it would be inefficient if we repeatedly attended to any given location. In fact, we exhibit **inhibition of return** (a reduced probability of returning to a region recently the focus of attention). Of theoretical importance is whether inhibition of return applies
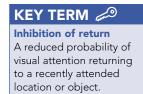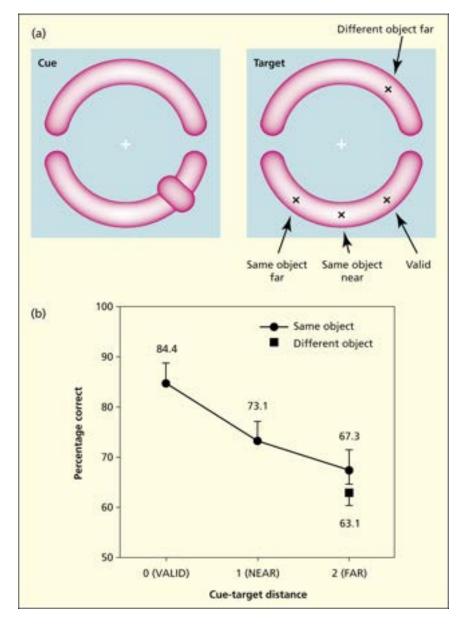
KEY TERM

Inhibition of return
A reduced probability of visual attention returning to a recently attended location or object.

Figure 5.4
(a) Possible target locations (same object far, same object near, valid, different object far) for a given cue. (b) Performance accuracy at the various target locations.

From Hollingworth et al. (2012). © 2011 American Psychological Association.

more to locations or objects. The evidence is mixed (see Chen, 2012). List and Robertson (2007) used Egly et al.'s (1994) task shown in Figure 5.4 and found location- or space-based inhibition of return was much stronger than object-based inhibition of return.

Theeuwes et al. (2014) found location- and object-based inhibition of return were *both* present at the same time. According to Theeuwes et al. (p. 2254), "If you direct your attention to a location in space, you will automatically direct attention to any object . . . present at that location, and vice versa."

There is considerable evidence of feature-based attention (see Bartsch et al., 2018, for a review). In their own research, Bartsch et al. addressed

the issue of whether feature-based attention to colour-defined targets is confined to the spatially attended region or whether it occurs across the entire visual field. They discovered the latter was the case.

Finally, Chen and Zelinsky (2019) argued it is important to study the allocation of attention under more naturalistic conditions than those typically used in research. In their study, observers engaged in free (unconstrained) viewing of natural scenes. Eye-fixation data suggested that attention initially selects regions of space. These regions may provide "the perceptual fragments from which objects are built" (p. 148).

Evaluation

Research on whether visual attention is object- or location-based have produced variable findings and so few definitive conclusions are possible. However, the relative importance of object-based and space- or location-based attention is *flexible*. For example, individual differences are important (Pilz et al., 2012). Note that visual attention can be both object-based and space-based at the same time.

What are the limitations of research in this area? First, most research apparently demonstrating that object-based attention is more important than space- or location-based attention has involved the use of spatial cues. Recent evidence (Donovan et al., 2017) suggests such cues may bias visual attention and that visual attention is *not* initially object-based in their absence.

Second, space-, object- and feature-based forms of attention often *interact* with each other to enhance object processing (Kravitz & Behrmann, 2011). However, we have as yet limited theoretical understanding of the mechanisms involved in such interactions.

Third, there is a need for more research assessing patterns of attention under naturalistic conditions. In recent research where observers view artificial stimuli while performing a specific task, it is unclear whether attentional processes resemble those when they engage in free viewing of natural scenes.

What happens to unattended or distracting stimuli?

Unsurprisingly, unattended visual stimuli receive less processing than attended ones. Martinez et al. (1999) compared event-related potentials (ERPs) to attended and unattended visual stimuli. The ERPs to unattended visual stimuli were comparable to those to attended ones 50–55 ms after stimulus onset. After that, however, the ERPs to attended stimuli were greater than those to unattended stimuli. Thus, selective attention influences all but the very early stages of processing.

As we have all discovered to our cost, it is often hard (or impossible) to ignore task-irrelevant stimuli. Below we consider factors determining whether task performance is adversely affected by distracting stimuli.

Load theory

Lavie's (2005, 2010) load theory has been an influential approach to understanding distraction effects. It distinguishes between perceptual

and cognitive load. Perceptual load refers to the perceptual demands of a current task. Cognitive load refers to the burden placed on the cognitive system by a current task (e.g., demands on working memory).

Tasks involving high perceptual load require nearly all our perceptual capacity whereas low-load tasks do not. With low-load tasks there are spare attentional resources, and so task-irrelevant stimuli are more likely to be processed. In contrast, tasks involving high cognitive load reduce our ability to use cognitive control to discriminate between target and distractor stimuli. Thus, high perceptual load is associated with *low* distractibility, whereas high cognitive load is associated with *high* distractibility.

Findings

There is much support for the hypothesis that high perceptual load reduces distraction effects. Forster and Lavie (2008) presented six letters in a circle and participants decided which target letter (X or N) was present. The five non-target letters resembled the target letter more closely in the high-load condition. On some trials a picture of a cartoon character (e.g., Spongebob Squarepants) was presented as a distractor outside the circle. Distractors interfered with task performance only under low-load conditions.
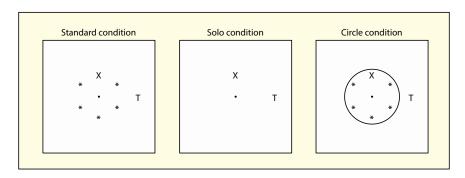
According to the theory, brain activation associated with distractors should be less when individuals are performing a task involving high perceptual load. This finding has been obtained with visual tasks and distractors (e.g., Schwartz et al., 2005) and also with auditory tasks and distractors (e.g., Sabri et al., 2013).

Why is low perceptual load associated with high distractibility? Biggs and Gibson (2018) argued this happens because observers generally adopt a broad attentional focus when perceptual load is low. They tested this hypothesis using three low-load conditions in which participants decided whether a target X or N was presented and a distractor letter was sometimes presented (see Figure 5.5). They argued that observers would adopt the smallest attentional focus in the circle condition and the largest attentional focus in the solo condition. As predicted, distractor interference was greatest in the solo condition and least in the circle condition. Thus, distraction effects depend strongly on size of attentional focus as well as perceptual load.

The hypothesis that distraction effects should be greater when cognitive or working memory load is high rather than low was tested by Burnham et al. (2014). As predicted, distraction effects on a visual search task were

Figure 5.5
Sample displays for three low perceptual load conditions in which the task required deciding whether a target X or N was presented. See text for further details.

From Biggs and Gibson (2018).

greater when participants performed another task placing high demands on the cognitive system.

Sörqvist et al. (2016) argued high cognitive load can *reduce* rather than increase distraction. They pointed out that cognitive load is typically associated with high levels of concentration and our everyday experience indicates high concentration generally reduces distractibility. As predicted, they found neural activation associated with auditory distractors was *reduced* when cognitive load on a visual task was high rather than low.

The effects of cognitive load on distraction are very variable. How can we explain this variability? Sörqvist et al. (2016) argued that an important factor is how easily distracting stimuli can be distinguished from task stimuli. When it is easy (e.g., task and distracting stimuli are in different modalities as in the Sörqvist et al., 2016, study), high cognitive load reduces distraction. In contrast, when it is hard to distinguish between task and distracting stimuli (e.g., they are similar and/or in the same modality), then high cognitive load increases distraction.

Load theory assumes the effects of perceptual and cognitive load are *independent*. However, Linnell and Caparos (2011) found perceptual and cognitive processes *interacted*: perceptual load only influenced attention as predicted when cognitive load was low. Thus, the effects of perceptual load are not "automatic" as assumed theoretically but instead depend on cognitive resources being available.

Evaluation

The distinction between perceptual and cognitive load has proved useful in predicting when distraction effects will be small or large. More specifically, the prediction that high perceptual load is associated with reduced distraction effects has received much empirical support. In applied research, load theory successfully predicts several aspects of drivers' attention and behaviour (Murphy & Greene, 2017). For example, drivers exposed to high perceptual load responded more slowly to hazards and drove less safely.

What are the theory's limitations? First, the terms "perceptual load" and "cognitive load" are vague, making it hard to test the theory (Murphy et al., 2016). Second, the assumption that perceptual and cognitive load have separate effects on attention is incorrect (Linnell & Caparos, 2011). Third, perceptual load and attentional breadth are often confounded. Fourth, the prediction that high cognitive load is associated with high distractibility has been disproved when task and distracting stimuli are easily distinguishable. Fifth, the theory de-emphasises several relevant factors including the salience or conspicuousness of distracting stimuli and the spatial distance between distracting and task stimuli (Murphy et al., 2016).

Major attention networks

As we saw in Chapter 1, many cognitive processes are associated with networks spread across relatively large areas of cortex rather than small, specific regions. With respect to attention, several theorists (e.g., Posner, 1980; Corbetta & Shulman, 2002) have argued there are two major networks.

One attention network is goal-directed or endogenous whereas the other is stimulus-driven or exogenous.

Posner's (1980) approach

Posner (1980) studied **covert attention** in which attention shifts to a given spatial location without an accompanying eye movement. In his research, people responded rapidly to a light. The light was preceded by a central cue (arrow pointing to the left or right) or a peripheral cue (brief illumination of a box outline). Most cues were *valid* (i.e., indicating where the target light would appear) but some were *invalid* (i.e., providing inaccurate information about the light's location).

Responses to the light were fastest to valid cues, intermediate to neutral cues (a central cross) and slowest to invalid cues. The findings were comparable for central and peripheral cues. When the cues were valid on only a small fraction of trials, they were ignored when they were central cues. However, they influenced performance when they were peripheral cues.

The above findings led Posner (1980) to distinguish between two attention systems:

(1) An *endogenous* system: it is controlled by the individual's intentions and is used when central cues are presented.
(2) An *exogenous* system: it automatically shifts attention and is involved when uninformative peripheral cues are presented. Stimuli that are salient or different from other stimuli (e.g., in colour) are most likely to be attended to using this system.

Corbetta and Shulman's (2002) approach

Corbetta and Shulman (2002) identified two attention systems that are involved in basic aspects of visual processing. First, there is a goal-directed or top-down system resembling Posner's endogenous system. This dorsal attention network consists of a fronto-parietal network including the intraparietal sulcus. It is influenced by expectations, knowledge and current goals. It is used when a cue *predicts* the location or other feature of a forthcoming visual stimulus.

Second, Corbetta and Shulman (2002) identified a stimulus-driven or bottom-up attention system resembling Posner's exogenous system. This is the ventral attention network and consists primarily of a right-hemisphere ventral fronto-parietal network. This system is used when an unexpected and potentially important stimulus (e.g., flames appearing under the door) occurs. Thus, it has a "circuit-breaking" function, meaning visual attention is redirected from its current focus. What stimuli trigger this circuit-breaking? According to Corbetta et al. (2008), non-task stimuli (i.e., distractors) closely resembling task stimuli are especially likely to activate the ventral attention network although salient or conspicuous stimuli also activate the same network.

Corbetta and Shulman (2011; see Figure 5.6) identified the brain areas associated with each network. Key areas within the dorsal attention network are as follows: superior parietal lobule (SPL), intraparietal sulcus
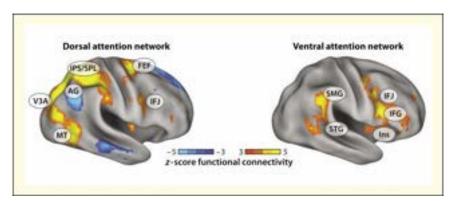

(IPS), inferior frontal junction (IFJ), frontal eye field (FEF), middle temporal area (MT) and V3A (a visual area). Key areas within the ventral attention network are as follows: inferior frontal junction (IFJ), inferior frontal gyrus (IFG), supramarginal gyrus (SMG), superior temporal gyrus (STG) and insula (Ins). The temporo-parietal junction also forms part of the ventral attention network.

The existence of two attention networks makes much sense. The goal-directed system (dorsal attention network) allows us to attend to stimuli directly relevant to our current goals. If we only had this system, however, our attentional processes would be dangerously inflexible. It is also important to have a stimulus-driven attentional system (ventral attention network) leading us to switch attention away from goal-relevant stimuli to unexpected threatening stimuli (e.g., a ferocious animal). More generally, the two attention networks typically *interact* effectively with each other.

Findings

Corbetta and Shulman (2002) supported their two-network model by carrying out meta-analyses of brain-imaging studies. In essence, they argued, brain areas most often activated when participants expect a stimulus that has not yet been presented form the dorsal attention network. In contrast, brain areas most often activated when individuals detect low-frequency targets form the ventral attention network.

Hahn et al. (2006) tested Corbetta and Shulman's (2002) theory by comparing patterns of brain activation when top-down and bottom-up processes were required. As predicted, there was little overlap between the brain areas associated with top-down and bottom-up processing. In addition, the brain areas involved in each type of processing corresponded reasonably well to those identified by Corbetta and Shulman.

Chica et al. (2013) reviewed research on the two attention systems and identified 15 differences between them. For example, stimulus-driven attention is faster than top-down attention and is more object-based. In addition, it is more resistant to interference from other peripheral cues once activated. The existence of so many differences strengthens the argument the two attentional systems are separate.

Considerable research evidence (mostly involving neuroimaging) indicates the dorsal and ventral attention systems are associated with distinct

neural circuits even during the resting state (Vossel et al., 2014). However, neuroimaging studies cannot establish that any given brain area is *necessarily* involved in stimulus-driven or goal-directed attention processes. Chica et al. (2011) provided relevant evidence by using transcranial magnetic stimulation (TMS; see Glossary) to interfere with processing in a given brain area. TMS applied to the right temporo-parietal junction impaired the functioning of the stimulus-driven system but not the top-down one.

In the same study, Chica et al. (2011) found TMS applied to the right intraparietal sulcus impaired the functioning of both attention systems. This provides evidence of the two attention systems working together.

Evidence from brain-damaged patients (discussed below, see pp. 196–200) is also relevant to establishing the brain areas necessarily involved in goal-directed or stimulus-driven attentional processes. Shomstein et al. (2010) had brain-damaged patients complete two tasks, one requiring stimulus-driven attentional processes whereas the other required top-down processes. Patients having greater problems with top-down attentional processing typically had brain damage to the superior parietal lobule (part of the dorsal attention network). In contrast, patients having greater problems with stimulus-driven attentional processing typically had brain damage to the temporo-parietal junction (part of the ventral attention network).

Wen et al. (2012) investigated *interactions* between the two visual attention systems. They assessed brain activation while participants responded to target stimuli in one visual field while ignoring all stimuli in the unattended visual field. There were two main findings. First, stronger causal influences of the top-down system on the stimulus-driven system led to *superior* performance on the task. This finding suggests the appearance of an object at the attended location caused the top-down attention system to suppress activity within the stimulus-driven system.

Second, stronger causal influences of the stimulus-driven system on the top-down system were associated with *impaired* task performance. This finding suggests activation within the stimulus-driven system produced by stimuli not in attentional focus disrupted the attentional set maintained by the top-down system.

Recent developments

Corbetta and Shulman's (2002) theoretical approach has been developed in recent years. Here we briefly consider three such developments. First, we now have a greater understanding of interactions between their two attention networks. Meyer et al. (2018) found stimulus-driven and goal-directed attention both activated frontal and parietal regions within the dorsal attention network, suggesting it has a pivotal role in integrating bottom-up and top-down processing.

Second, previous research reviewed by Corbetta and Shulman (2002) indicated the dorsal attention network is active immediately prior to the presentation of an anticipated visual stimulus. However, this research did not indicate how long this attention network remained active. Meehan et al. (2017) addressed this issue and discovered that top-down influences associated within the dorsal attention network persisted over a relatively long time period.

Third, brain networks relevant to attention additional to those within Corbetta and Shulman's (2002) theory have been identified (Sylvester et al., 2012). One such network is the cingulo-opercular network including the anterior insula/operculum and dorsal anterior cingulate cortex (dACC; see Figure 5.7). This network is associated with non-selective attention or alertness (Coste & Kleinschmidt, 2016).

Another additional network is the default mode network including the posterior cingulate cortex (PCC), the lateral parietal cortex (LP), the inferior temporal cortex (IT), the medial prefrontal cortex (MPF) and the subgenual anterior cingulate cortex (sgACC). The **default mode network** is activated during *internally* focused cognitive processes (e.g., mind-wandering; imagining the future). What is the relevance of this network to attention? In essence, performance on tasks requiring *externally* focused attention is often enhanced if the default mode network is deactivated (Amer et al., 2016a).

Finally, there is the fronto-parietal network (Dosenbach et al., 2008), which includes the anterior dorsolateral prefrontal cortex (aDLPFC), the

KEY TERM

Default mode network
A network of brain regions that is active "by default" when an individual is not involved in a current task; it is associated with internal processes including mind-wandering, remembering the past and imagining the future.
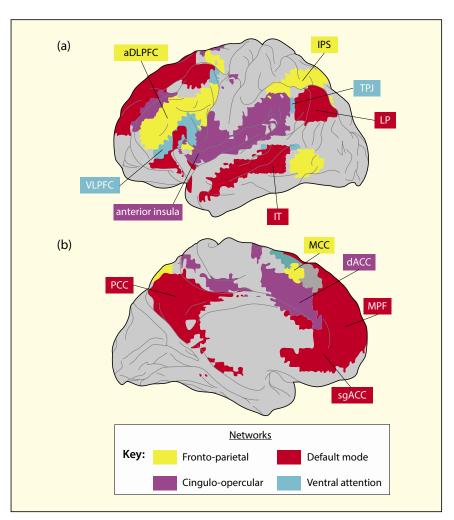
Figure 5.7
This is part of a theoretical approach based on several functional networks of relevance to attention: the four networks shown (fronto-parietal; default mode; cingulo-opercular; and ventral attention) are all discussed fully in the text.
Sylvester et al., 2012, p. 528. Reprinted with permission of Elsevier.

middle cingulate cortex (MCC) and the intraparietal sulcus (IPS). It is associated with top-down attentional and cognitive control.

Evaluation

The theoretical approach proposed by Corbetta and Shulman (2002) has several successes to its credit. First, there is convincing evidence for somewhat separate stimulus-driven and top-down attention systems, each with its own brain network. Second, research using transcranial magnetic stimulation suggests major brain areas within each attention system play a *causal* role in attentional processes. Third, some interactions between the two networks have been identified. Fourth, research on brain-damaged patients supports the theoretical approach (see next section, pp. 196–200).

What are the limitations of this theoretical approach? First, the precise brain areas associated with each attentional system have not been clearly identified. Second, there is more commonality (especially within the parietal lobe) in the brain areas associated with the two attention networks than assumed theoretically by Corbetta and Shulman (2002). Third, there are additional attention-related brain networks not included within the original theory. Fourth, much remains to be discovered about how different attention systems interact.

DISORDERS OF VISUAL ATTENTION

Here we consider two important attentional disorders in brain-damaged individuals: neglect and extinction. **Neglect** (or spatial neglect) involves a lack of awareness of stimuli presented to the side of space on the opposite side to the brain damage (the contralesional side). This occurs because information from the left side of the visual field proceeds to the right hemisphere.

Most neglect patients have damage in the right hemisphere and so lack awareness of stimuli on the left side of the visual field: *space-based* or *egocentric neglect*. For example, patients crossing out targets presented to their left or right side (cancellation task) cross out more of those presented to the right. When instructed to mark the centre of a horizontal line (line bisection task), patients put it to the right of the centre. Note that the right hemisphere is dominant in spatial attention in healthy individuals – they exhibit **pseudo-neglect**, in which the left side of visual space is favoured (Friedrich et al., 2018).

There is also *object-centred* or *allocentric neglect* involving a lack of awareness of the left side of *object*s (see Figure 5.8). Patients with right-hemisphere damage typically draw the right side of all figures in a multi-object scene but neglect their left side in the left and right visual fields (Gainotti & Ciaraffa, 2013).

Do allocentric and egocentric neglect reflect a *single* disorder or *separate* disorders? Rorden et al. (2012) obtained two findings supporting the single disorder explanation. First, the correlation between the extent of each form of neglect across 33 patients was +.80. Second, similar brain regions were associated with each type of neglect. However, Pedrazzini et al. (2017) found damage to the intraparietal sulcus was more associated
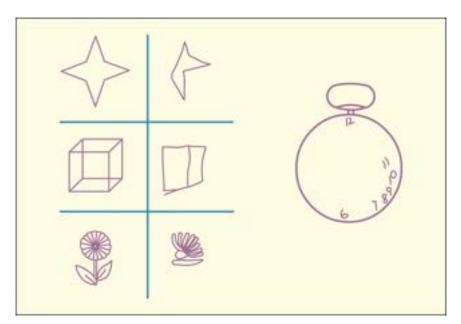
with allocentric than egocentric neglect, whereas the opposite was the case with damage to the temporo-parietal junction.

Extinction is often found in neglect patients. **Extinction** involves a failure to detect a stimulus presented to the side opposite the brain damage when a second stimulus is presented to the same side as the brain damage. Extinction and neglect are closely related but separate deficits (de Haan et al., 2012). We will focus mostly on neglect because it has attracted much more research.

Which brain areas are damaged in neglect patients? Neglect is a heterogeneous condition and the brain areas damaged vary considerably across patients. In a meta-analysis, Molenberghs et al. (2012) found the main areas damaged in neglect patients are in the right hemisphere and include the superior temporal gyrus, the inferior frontal gyrus, the insula, the supramarginal gyrus and the angular gyrus (gyrus means ridge). Nearly all these areas are within the stimulus-driven or ventral attention network (see Figure 5.6) suggesting brain *networks* are damaged rather than simply specific brain *areas* (Corbetta & Shulman, 2011).

We also need to consider functional connectivity (correlated brain activity between brain regions). Baldassarre et al. (2014, 2016) discovered widespread disruption of functional connectivity between the hemispheres in neglect patients. This disruption did not involve the bottom-up and top-down attention networks. Of importance, recovery from attention deficits in neglect patients was associated with improvements in functional connectivity in bottom-up and top-down attention networks (Ramsey et al., 2016).

The right-hemisphere temporo-parietal junction and intraparietal sulcus are typically damaged in extinction patients (de Haan et al., 2012). When transcranial magnetic stimulation is applied to these areas to interfere with processing, extinction-like behaviour results (de Haan et al., 2012). Dugué et al. (2018) confirmed the importance of the temporo-parietal junction

Interactive feature:
Primal Pictures'
3D atlas of the brain

KEY TERM

Extinction
A disorder of visual attention in which a stimulus presented to the side opposite the brain damage is not detected when another stimulus is presented at the same time to the side of the brain damage.

(part of the ventral attention network) in control of spatial attention in a neuroimaging study on healthy individuals. However, its subregions varied in terms of their involvement in voluntary and involuntary attention shifts.

Conscious awareness and processing

Neglect patients generally report no conscious awareness of stimuli presented to the left visual field. However, that does not necessarily mean those stimuli are not processed. Vuilleumier et al. (2002b) presented extinction patients with two pictures at the same time, one to each visual field. The patients showed very little memory for left-field stimuli. Then the patients identified degraded pictures. There was a facilitation effect for left-field pictures indicating they had been processed.

Vuilleumier et al. (2002a) presented GK, a male patient with neglect and extinction, with fearful faces. He showed increased activation in the amygdala (associated with emotional responses) whether or not these faces were consciously perceived. This is explicable given there is a processing route from the retina to the amygdala bypassing the cortex (Diano et al., 2017).

Sarri et al. (2010) found extinction patients had no awareness of left-field stimuli. However, these stimuli were associated with activation in early visual processing areas, indicating they received some processing.

Processing in neglect and extinction has been investigated using event-related potentials. Di Russo et al. (2008) focused on the processing of left-field stimuli not consciously perceived by neglect patients. Early processing of these stimuli was comparable to that of healthy controls with only later processing being disrupted. Lasaponara et al. (2018) obtained similar findings in neglect patients. In healthy individuals, the presentation of left-field targets inhibits processing of right-field space. This was less the case in neglect patients, which helps to explain their lack of conscious perception of left-field stimuli.

Theoretical considerations

Corbetta and Shulman (2011) discussed neglect in the context of their two-system theory (discussed earlier, see pp. 192–196). In essence, the bottom-up ventral attention network is damaged. Strong support for this assumption was reported by Toba et al. (2018a) who found in 25 neglect patients that impaired performance on tests of neglect was associated with damage to parts of the ventral attention network (e.g., angular gyrus; supramarginal gyrus). Since the right hemisphere is dominant in the ventral attention network, neglect patients typically have damage in that hemisphere. Of importance, Corbetta and Shulman (2011) also assumed that damage to the ventral network impairs the functioning of the goal-directed dorsal attention network (even though not itself damaged).

How does the damaged ventral attention network impair the dorsal attention network's functioning? The two attention networks *interact* and so damage to the ventral network inevitably affects the functioning of the dorsal network. More specifically, damage to the ventral attention network "impairs non-spatial [across the entire visual field] functions, hypoactivates

[reduces activation in] the right hemisphere, and unbalances the activity of the dorsal attention network" (Corbetta & Shulman, 2011, p. 592).

de Haan et al. (2012) proposed a theory of extinction based on two major assumptions:

(1) "Extinction is a consequence of biased *competition* for attention between the ipsilesional [right-field] and contralesional [left-field] target stimuli" (p. 1048);
(2) Extinction patients have much reduced attentional capacity so often only one target [the right-field one] can be detected.

Findings

According to Corbetta and Shulman (2011), the dorsal attention network in neglect patients functions poorly because of reduced activation in the right hemisphere and associated reduced alertness and attentional resources. Thus, increasing patients' general alertness should enhance their detection of left-field visual targets. Robertson et al. (1998) found the slower detection of left visual field stimuli compared to those in the right visual field was no longer present when warning sounds were used to increase alertness.

Bonato and Cutini (2016) compared neglect patients' ability to detect visual targets with (or without) a second, attentionally demanding task. Detection rates were high for targets presented to the right visual field in both conditions. In contrast, patients detected only approximately 50% as many targets in the left visual field as the right when performing another task. Thus, neglect patients have limited attentional resources.

Corbetta and Shulman (2011) assumed neglect patients have an essentially intact dorsal attention network. Accordingly, neglect patients might use that network effectively if steps were taken to facilitate its use. Duncan et al. (1999) presented arrays of letters and neglect patients recalled only those in a pre-specified colour (the dorsal attention network could be used to select the appropriate letters). Neglect patients resembled healthy controls in showing equal recall of letters presented to each side of visual space.

The two attention networks typically work closely together. Bays et al. (2010) studied neglect patients. They used eye movements during a visual search to assess patients' problems with top-down and stimulus-driven attentional processes. Both types of attentional processes were equally impaired (as predicted by Corbetta and Shulman, 2011). Of most importance, there was a remarkably high correlation of +.98 between these two types of attentional deficit.

Toba et al. (2018b) identified two reasons for the failure of neglect patients to detect left-field stimuli:

(1) a "magnetic" attraction of attention (i.e., right-field stimuli immediately capture attention).
(2) impaired spatial working memory making it hard for patients to keep track of the locations of stimuli.

Both reasons were equally applicable to most patients. However, the first reason was dominant in 12% of patients and the second reason in

24% of patients. Accordingly, Toba et al. argued we should develop *multi-component* models of visual neglect to account for such individual differences.

We turn now to extinction patients. According to de Haan et al. (2012), extinction occurs because of biased competition between stimuli. If two stimuli could be *integrated*, that might minimise competition and so reduce extinction. Riddoch et al. (2006) tested this prediction by presenting objects used together often (e.g., *wine bottle* and *wine glass*) or never used together (e.g., *wine bottle* and *ball*). Extinction patients identified both objects more often in the former condition than the latter (65% vs 40%, respectively).

The biased competition hypothesis has been tested in other ways. We can impair attentional processes in the intact left hemisphere by applying transcranial magnetic stimulation to it. This should reduce competition from the left hemisphere in extinction patients and thus reduce extinction. Some findings are consistent with this prediction (Oliveri & Caltagirone, 2006).

de Haan et al. (2012) also identified reduced attentional capacity as a factor causing extinction. Bonato et al. (2010) studied extinction with or without the addition of a second, attentionally demanding task. As predicted, extinction patients showed a substantial increase in the extinction rate (from 18% to over 80%) with this additional task.

Overall evaluation

Research has produced several important findings. First, neglect and extinction patients can process unattended visual stimuli in the absence of conscious awareness of those stimuli. Second, most neglect patients have damage to the ventral attention network leading to impaired functioning of the undamaged dorsal attention network. Third, extinction occurs because of biased competition for attention and reduced attentional capacity.

What are the limitations of research in this area? First, it is hard to produce theoretical accounts applicable to all neglect or extinction patients because the precise symptoms and regions of brain damage vary considerably across patients. Second, neglect patients vary in their precise processing deficits (e.g., Toba et al., 2018b), but this has been de-emphasised in most theories. Third, the precise relationship between neglect and extinction remains unclear. Fourth, the dorsal and ventral networks generally interact but the extent of their interactions remains to be determined.

VISUAL SEARCH

We spend much time searching for various objects (e.g., a friend in a crowd). The processes involved have been studied in research on **visual search** where a specified target is detected as rapidly as possible. Initially, we consider an important real-world situation where visual search can be literally a matter of life-or-death: airport security checks. After that, we consider an early very influential theory of visual search before discussing more recent theoretical and empirical developments.

IN THE REAL WORLD: AIRPORT SECURITY CHECKS 🌐

Airport security checks have become more thorough since 9/11. When your luggage is x-rayed, an airport security screener searches for illegal and dangerous items (see Figure 5.9). Screeners are well trained but mistakes sometimes occur.

Figure 5.9
Each bag contains one illegal item. From left to right: a large bottle; a dynamite stick; and a gun part.
From Mitroff & Biggs (2014).

There are two major reasons it is often hard for airport security screeners to detect dangerous items. First, illegal and dangerous items are (thankfully!) present in only a minute fraction of passengers' luggage. This rarity of targets makes it hard for airport security screeners to detect them. Mitroff and Biggs (2014) asked observers to detect illegal items in bags (see Figure 5.9). The detection rate was only 27% when targets appeared under 0.15% of the time: they termed this the "ultra rare item effect". In contrast, the detection rate was 92% when targets appeared more than 1% of the time.

Peltier and Becker (2016) tested two explanations for the reduced detection rate with rare targets: (1) a reduced probability that the target is fixated (selection error); and (2) increased caution about reporting targets because they are so unexpected (identification error). There was evidence for both explanations. However, most detection failures were selection errors (see Figure 5.10).
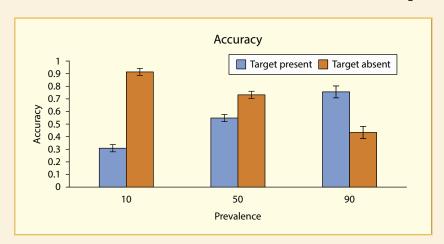

Figure 5.10
Frequency of selection and identification errors when targets were present on 10%, 50% or 90% of trials.
From Peltier and Becker (2016).

Second, security screeners search for numerous different objects. This increases search difficulty. Menneer et al. (2009) found target detection was worse when screeners searched for two categories of objects (metal threats and improvised explosive devices) rather than one.

How can we increase the efficiency of security screening? First, we can exploit individual differences in the ability to detect targets. Rusconi et al. (2015) found individuals scoring high on a questionnaire measure of attention to detail had superior target-detection performance than low scorers.

Second, airport security screeners can find it hard to distinguish between targets (i.e., dangerous items) and similar-looking non-targets. Geng et al. (2017) found that observers whose training included non-targets resembling targets learned to develop increasingly precise internal target representations. Such representations can improve the speed and accuracy of security screening.

Third, the low detection rate when targets are very rare can be addressed. Threat image projection (TIP) can be used to project fictional threat items into x-ray images of luggage to increase the apparent frequency of targets. When screeners are presented with TIPs plus feedback when they miss them, screening performance improves considerably (Hofer & Schwaninger, 2005). In similar fashion, Schwark et al. (2012) found providing false feedback to screeners to indicate they had missed rare targets reduced their cautiousness about reporting targets and improved their performance.

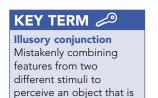## Feature integration theory

Feature integration theory was proposed by Treisman and Gelade (1980) and subsequently updated and modified (e.g., Treisman, 1998). According to the theory, we need to distinguish between object features (e.g., colour; size; line orientation) and the objects themselves. There are two processing stages:

(1) Basic visual features are processed rapidly and pre-attentively in *parallel* across the visual scene.
(2) Stage (1) is followed by a slower *serial* process with focused attention providing the "glue" to form objects from the available features (e.g., an object that is round and has an orange colour is perceived as an orange). In the absence of focused attention, features from different objects may be combined randomly producing an **illusory conjunction**.

It follows from the above assumptions that targets defined by a *single* feature (e.g., a blue letter or an S) should be detected rapidly and in parallel. In contrast, targets defined by a *conjunction* or combination of features (e.g., a green letter T) should require focused attention and so should be slower to detect. Treisman and Gelade (1980) tested these predictions using both types of targets; the display size was 1–30 items and a target was present or absent.

As predicted, response was rapid and there was very little effect of display size when the target was defined by a single feature: these findings suggest parallel processing (see Figure 5.11). Response was slower and was strongly influenced by display size when the target was defined by a conjunction of features: these findings suggest there was serial processing.

According to the theory, lack of focused attention can produce illusory conjunctions based on random combinations of features. Friedman-Hill

KEY TERM 🔑

Illusory conjunction
Mistakenly combining features from two different stimuli to perceive an object that is not present.
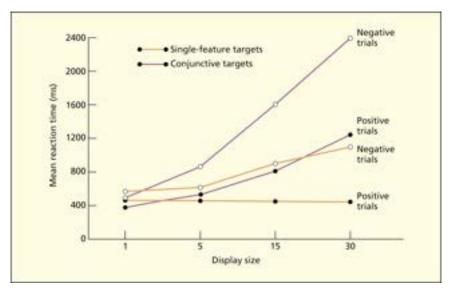
Figure 5.11
Performance speed on a detection task as a function of target definition (conjunctive vs single feature) and display size.

Adapted from Treisman and Gelade (1980).

et al. (1995) studied a brain-damaged patient (RM) having problems with the accurate location of visual stimuli. This patient produced many illusory conjunctions combining the shape of one stimulus with the colour of another.

Limitations

What are the theory's limitations? First, Duncan and Humphreys (1989, 1992) identified two factors not included within feature integration theory:

(1) When distractors are very similar to each other, visual search is faster because it is easier to identify them as distractors.
(2) The number of distractors has a strong effect on search time to detect even targets defined by a single feature when targets resemble distractors.

Second, Treisman and Gelade (1980) estimated the search time with conjunctive targets was approximately 60 ms per item and argued this represented the time taken for focal attention to process each item. However, research with other paradigms indicates it takes approximately 250 ms for attention indexed by eye movements to move from one location to another. Thus, it is improbable focal attention plays the key role assumed within the theory.

Third, the theory assumes visual search is often item-by-item. However, the information contained within most visual scenes cannot be divided up into "items" and so the theory is of limited applicability. Such considerations led Hulleman and Olivers (2017) to produce an article entitled "The impending demise of the item in visual search".

Fourth, visual search involves parallel processing much more than implied by the theory. For example, Thornton and Gilden (2007) used 29 different visual tasks and found 72% apparently involved parallel

processing. We can explain such findings by assuming that each eye fixation permits considerable parallel processing using information available in peripheral vision (discussed below, see pp. 206–208).

Fifth, the theory assumes that the early stages of visual search are entirely feature-based. However, recent research using event-related potentials indicates that object-based processing can occur much faster than predicted by feature integration theory (e.g., Berggren & Eimer, 2018).

Sixth, the theory assumes visual search is essentially *random*. This assumption is wrong with respect to the real world – we typically use our knowledge of where a target object is likely to be located when searching for it (see below).

Dual-path model

In most of the research discussed so far, the target appeared at a *random* location within the visual display. This is radically different from the real world. Suppose you are outside looking for your missing cat. Your visual search would be very selective – you would ignore the sky and focus mostly on the ground (and perhaps the trees). Thus, your search would involve *top-down processes* based on your knowledge of where cats are most likely to be found.

Ehinger et al. (2009) studied top-down processes in visual search by recording eye fixations of observers searching for a person in 900 real-world outdoor scenes. Observers typically fixated plausible locations (e.g., pavements) and ignored implausible ones (e.g., sky; trees; see Figure 5.12). Observers also fixated locations differing considerably from neighbouring locations and areas containing visual features resembling those of a human figure.

How can we reconcile Ehinger et al.'s (2009) findings with those discussed earlier? Wolfe et al. (2011) proposed a dual-path model (see Figure 5.13). There is a *selective* pathway of limited capacity (indicated by the bottleneck) with objects being selected individually for recognition.
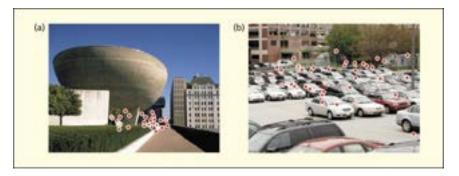
Figure 5.12

The first three eye fixations made by observers searching for pedestrians. As can be seen, the great majority of their fixations were on regions in which pedestrians would most likely be found. Observers' fixations were much more like each other in the left-hand photo than in the right-hand one, because there were fewer likely regions in the left-hand one.

From Ehinger et al. (2009). Reprinted with permission from Taylor & Francis.
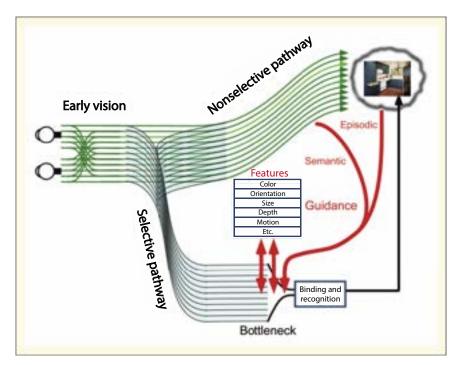
Figure 5.13

A two-pathway model of visual search. The selective pathway is capacity limited and can bind stimulus features and recognise objects. The non-selective pathway processes the gist of scenes. Selective and non-selective processing occur in parallel to produce effective visual search.

From Wolfe et al. (2011). Reprinted with permission from Elsevier.

This pathway has been the focus of most research until recently. There is also a *non-selective* pathway in which the "gist" of a scene is processed. Such processing can then guide processing within the selective pathway (represented by the arrow labelled "guidance"). This pathway allows us to utilise our stored environmental knowledge and so is of great value in the real world.

Findings

Wolfe et al. (2011) compared visual searches for objects presented within a scene setting or at random locations. As predicted, search rate per item was much faster in the scene setting (10 ms vs 40 ms, respectively). Võ and Wolfe (2012) explained that finding in terms of "functional set size" – searching in scenes is efficient because most regions can be ignored. As predicted, Võ and Wolfe found 80% of each scene was rarely fixated.

Kaiser and Cichy (2018) presented observers with objects typically located in the upper (e.g., *aeroplane*; *hat*) or lower (e.g., *carpet*; *shoe*) visual field. These objects were presented in their typical or atypical location (e.g., *hat* in the lower visual field). Observers had to indicate whether an object presented very briefly was located in the upper or lower visual field. Observers' performance was better when objects appeared in their typical location because of their extensive knowledge of where objects are generally located.

Chukoskie et al. (2013) found observers can easily *learn* where targets are located. An invisible target was presented at random locations on a blank screen and observers were provided with feedback. There was a strong learning effect – fixations rapidly shifted from being fairly

Fovea
A small area within the retina in the centre in the field of vision where visual acuity is greatest.

random to being focused on the area within which the target might be present.

Ehinger et al.'s (2009) findings (discussed earlier, see p. 204) suggested that scene gist or context can be used to enhance the efficiency of visual search. Katti et al. (2017) presented scenes very briefly (83 ms) followed by a mask. Observers were given the task of detecting a person or a car and performed very accurately (over 90%) and rapidly. Katti et al. confirmed that scene gist or context influenced performance. However, performance was influenced more strongly by features of the target object – the more key features of an object were visible, the faster it was detected.

What is the take-home message from the above study? The efficiency of visual search with real-world scenes is more complex than implied by Ehinger et al. (2009). More specifically, observers may rapidly fixate on the area close to a target person because they are using scene gist or because they rapidly process features of the person (e.g., wearing clothes).

Evaluation

Our knowledge of likely (and unlikely) locations for any given object in a scene influences visual search in the real world. This is fully acknowledged in the dual-path model. There is also support for the notion that scene knowledge facilitates visual search by reducing functional set size.

What are the model's limitations? First, how we use gist knowledge of a scene very rapidly to reduce the search area remains unclear. Second, there is insufficient focus on the learning processes that can greatly facilitate visual search – the effects of such processes can be seen in the very rapid and accurate detection of target information by experts in several domains (see Chapter 11).

Third, it is important not to exaggerate the importance of scene gist or context in influencing the efficiency of visual search. Features of the target object can influence visual search more than scene gist (Katti et al., 2017).

Fourth, the assumption that items are processed individually within the selective pathway is typically mistaken. As we will see shortly, visual search often depends on parallel processes within peripheral vision and such processes are not considered within the model.

Attention vs perception: texture tiling model

Several theories (e.g., Treisman & Gelade, 1980) have assumed that individual items are the crucial units in visual search. Such theories have also often assumed that slow visual search depends mostly on the limitations of focused attention. A plausible implication of these assumptions is that slow visual search depends mostly on foveal vision (the **fovea** is a small area of maximal visual acuity in the retina).

Both the above assumptions have been challenged recently. At the risk of oversimplification, full understanding of visual search requires less emphasis on *attention* and more on *perception*. According to Rosenholtz (2016), peripheral (non-foveal) vision is of crucial importance. Acuity decreases as we move away from the fovea to the periphery of vision, but much less than often assumed. You can demonstrate this by holding out

your thumb and fixating the nail. Foveal vision only covers the nail so the great majority of what you can see is in peripheral vision.

We can also compare the value of foveal and peripheral vision by considering individuals with impaired eyesight. Those with severely impaired peripheral vision (e.g., due to glaucoma) had greater problems with mobility (e.g., number of falls; ability to drive) than those who lack foveal vision (due to macular degeneration) (Rosenholtz, 2016). Individuals with severely impaired central or foveal vision performed almost as well as healthy controls at detecting target objects in coloured scenes (75% vs 79%, respectively) (Thibaut et al., 2018).

If visual search depends heavily on peripheral vision, what predictions can we make? First, if each fixation provides observers with a considerable amount of information about several objects, visual search will typically involve parallel rather than serial processing. Second, we need to consider limitations of peripheral vision (e.g., visual acuity is less in peripheral than foveal vision). However, a more important limitation concerns **visual crowding** – a reduced ability to recognise objects or other stimuli because of irrelevant neighbouring objects or stimuli (clutter). Visual crowding impairs peripheral vision to a much greater extent than foveal vision.

Rosenholtz et al. (2012) proposed the texture tiling model based on the assumption peripheral vision is of crucial importance in visual search. More specifically, processing in peripheral vision can cause adjacent stimuli to tile (join together) to form an apparent target, thus increasing the difficulty of visual search. Below we consider findings relevant to this model.

Findings

As mentioned earlier (p. 203), Thornton and Gilden (2007) found almost three-quarters of the visual tasks they studied involved parallel processing. This is entirely consistent with the emphasis on parallel processing in the model.

Direct evidence for the importance of peripheral vision to visual search was reported by Young and Hulleman (2013). They manipulated the visible area around the fixation point making it small, medium or large. As predicted by the model, visual search performance was worst when the visible area was small (so only one item could be processed per fixation). Overall, visual search was almost parallel when the visible area was large but serial when it was small.

Chang and Rosenholtz (2016) used various search tasks. According to feature integration theory, both tasks shown in Figure 5.14 should be comparably hard because the target and distractors share features. In contrast, the texture tiling model predicts the task on the right should be harder because adjacent distractors seen in peripheral vision can more easily tile (join together) to form an apparent T. The findings from these tasks (and several others) supported the texture tiling model but were inconsistent with feature integration theory.
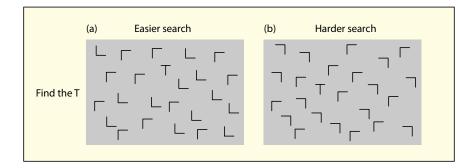
Finally, Hulleman and Olivers (2017) produced a model of visual search consistent with the texture tiling model. According to this model, each eye fixation lasts 250 ms, during which information from foveal and peripheral vision is extracted in parallel. They also assumed that the area

Figure 5.14
The target (T) is easier to find in the display on the left than the one on the right.

From Chang and Rosenholtz (2016).

around the fixation point within which a target can generally be detected is smaller when the visual search task is difficult (e.g., because target discriminability is low).

A key prediction from Hulleman and Olivers' (2017) model is that the main reason why search times are longer with more difficult search tasks is because more eye fixations are required than with easier tasks. A computer simulation based on these assumptions produced search times very similar to those obtained in experimental studies.

Evaluation

What are the strengths of the texture tiling model? First, the information available in peripheral vision is much more important in visual search than assumed previously. The model explains how observers make use of the information available in peripheral vision.

Second, the model explains why parallel processing is so prevalent in visual search – it reflects directly parallel processing within peripheral vision. Third, there is accumulating evidence that search times are generally directly related to the number of eye fixations.

Fourth, an approach based on eye fixations and peripheral vision can potentially explain findings from *all* visual search paradigms, including complex visual scenes and item displays. Such an approach thus has more general applicability than feature integration theory.

What are the model's limitations? First, as Chang and Rosenholtz (2016) admitted, it needs further development to account fully for visual search performance. For example, it does not predict search times with precision. In addition, it does not specify the criteria used by observers to decide no target is present.

Second, visual search is typically much faster for experts than non-experts in their domain of expertise (e.g., medical experts examining mammograms) (see Chapter 11). The texture tiling model does not identify clearly the processes allowing experts to make very efficient use of peripheral information.

CROSS-MODAL EFFECTS

Nearly all the research discussed so far is limited in that the visual (or auditory) modality was studied on its own. We might try to justify this approach by assuming attentional processes in each sensory modality operate

independently from those in other modalities. However, that assumption is incorrect. In the real world, we often coordinate information from two or more sense modalities at the same time (**cross-modal attention**). An example is lip reading, where we use visual information about a speaker's lip movements to facilitate our understanding of what they are saying (see Chapter 9).

Suppose we present participants with two streams of lights (as was done by Eimer and Schröger, 1998), with one stream being presented to the left and the other to the right. At the same time, we present participants with two streams of sounds (one to each side). In one condition, participants detect deviant *visual* events (e.g., longer than usual stimuli) presented to one side only. In the other condition, participants detect deviant *auditory* events in only one stream.

Event-related potentials were recorded to assess the allocation of attention. Unsurprisingly, Eimer and Schröger (1998) found ERPs to deviant stimuli in the *relevant* modality were greater to stimuli presented on the to-be-attended side than the to-be-ignored side. Thus, participants allocated attention as instructed.

Of more interest is what happened to the allocation of attention in the *irrelevant* modality. Suppose participants detected *visual* targets on the left side. In that case, ERPs to deviant *auditory* stimuli were greater on the left side than the right. This is a cross-modal effect: the voluntary or endogenous allocation of *visual* attention also affected the allocation of *auditory* attention. Similarly, when participants detected *auditory* targets on one side, ERPs to deviant *visual* stimuli on the same side were greater than ERPs to those on the opposite side. Thus, the allocation of *auditory* attention also influenced the allocation of *visual* attention.

Ventriloquism effect

What happens when there is a conflict between simultaneous visual and auditory stimuli? We will focus on the **ventriloquism effect** in which sounds are misperceived as coming from their apparent visual source. Ventriloquists (at least good ones!) speak without moving their lips while manipulating a dummy's mouth movements. It seems as if the dummy is speaking. Something similar happens at the movies. The actors' lips move on the screen but their voices come from loudspeakers beside the screen. Nevertheless, we hear those voices coming from their mouths.

Certain conditions must be satisfied for the ventriloquism effect to occur (Recanzone & Sutter, 2008). First, the visual and auditory stimuli must occur close together in time. Second, the sound must match *expectations* created by the visual stimulus (e.g., high-pitched sound coming from a small object). Third, the sources of the visual and auditory stimuli should be close together spatially. More generally, the ventriloquism effect reflects the unity assumption (the assumption that two or more sensory cues come from the same object: Chen & Spence, 2017).

The ventriloquism effect exemplifies *visual dominance* (visual information dominating perception). Further evidence comes from the Colavita effect (Colavita, 1974): participants instructed to respond to all stimuli respond more often to visual than simultaneous auditory stimuli (Spence et al., 2011).

When during processing is visual spatial information integrated with auditory information? Shrem et al. (2017) found that misleading visual information about the location of an auditory stimulus influenced the processing of the auditory stimulus approximately 200 ms after stimulus onset. The finding that this effect is still present even when participants are aware of the spatial discrepancy between the visual and auditory input suggests it occurs relatively "automatically".

However, the ventriloquism effect is smaller when participants had previously heard syllables spoken in a fearful voice (Maiworm et al., 2012). This suggests the effect is not entirely "automatic" but is reduced when the relevance of the auditory channel is increased.

Why does vision capture sound in the ventriloquism effect? The visual modality typically provides more precise information about spatial location. However, when visual stimuli are severely blurred and poorly localised, sound captures vision (Alais & Burr, 2004). Thus, we combine visual and auditory information effectively by attaching more weight to the more informative sense modality.

Temporal ventriloquism

The above explanation for the ventriloquist illusion is a development of the modality appropriateness and precision hypothesis (Welch & Warren, 1980). According to this hypothesis, when conflicting information is presented in two or more modalities, the modality having the greatest acuity generally dominates. This hypothesis predicts the existence of another illusion. The auditory modality is typically more precise than the visual modality at discriminating temporal relations. As a result, judgements about the temporal onset of visual stimuli might be biased by auditory stimuli presented very shortly beforehand or afterwards. This is the **temporal ventriloquism effect**.
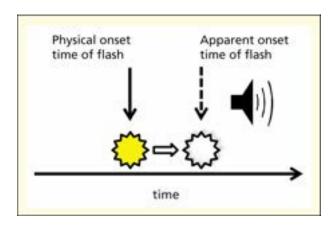

Figure 5.15
An example of temporal ventriloquism in which the apparent time of onset of a flash is shifted towards that of a sound presented at a slightly different timing from the flash.

From Chen and Vroomen (2013). Reprinted with permission from Springer.

Research on temporal ventriloquism was reviewed by Chen and Spence (2017). A simple example is when the apparent onset of a flash is shifted towards an abrupt sound presented slightly asynchronously (see Figure 5.15). Other research has found that the apparent duration of visual stimuli can be distorted by asynchronous auditory stimuli.

We need to consider the temporal ventriloquism effect in the context of the unity assumption. This is the assumption that "two or more uni-sensory cues belong together (i.e., that they come from the same object or event)" (Chen & Spence, 2017, p. 1). Chen and Spence discussed findings showing that the unity assumption generally (but not always) enhances the temporal ventriloquism effect.

Orchard-Mills et al. (2016) extended research by using two visual stimuli (one

IN THE REAL WORLD: WARNING SIGNALS PROMOTE SAFE DRIVING

Front-to-rear-end collisions cause 25% of road accidents with driver inattention the most common cause (Spence, 2012). Thus, it is important to devise effective warning signals to enhance driver attention and reduce collisions. Warning signals might be especially useful if they were *informative* (i.e., indicating the nature of the danger). However, informative warning signals requiring time-consuming cognitive processing might be counterproductive.

Ho and Spence (2005) considered drivers' reaction times when braking to avoid a car in front or accelerating to avoid a speeding car behind. An auditory warning signal (car horn) came from the same direction as the critical visual event on 80% or 50% of trials. Braking times were faster when the sound and critical visual event were from the same direction. The greater beneficial effects of auditory signals when predictive rather than non-predictive suggests the involvement of **endogenous spatial attention** (controlled by the individual's intentions). Auditory stimuli also influenced visual attention even when *non*-predictive: this probably involved **exogenous spatial attention** ("automatic" allocation of attention).

Gray (2011) studied braking times to avoid a collision with the car in front when drivers heard auditory warning signals increasing in intensity as the time to collision reduced. These signals are known as looming sounds. The most effective condition was the one where the rate of increase in the intensity of the auditory signal was the fastest because it implied the time to collision was the least. Lahmer et al. (2018) found evidence that looming sounds are effective because they are consistent with the visual experience of an approaching collision.

Vibrotactile signals produce the perception of vibration through touch. Gray et al. (2014) studied the effects of such signals on speed of braking to avoid a collision. Signals were presented at three sites on the abdomen arranged vertically. In the most effective condition, successive signals moved towards the driver's head at an increasing rate reflecting the speed they were approaching the car in front. Braking time was 250 ms faster in this condition than a no-warning control condition, probably because it was highly informative.

Ahtamad et al. (2016) compared the effectiveness of three vibrotactile warning signals delivered to the back on braking times to avoid a collision with the car in front: (1) *expanding* (centre of back followed by areas to left and right); (2) *contracting* (areas to left and right followed by the centre of the back); (3) *static* (centre of the back + areas to left and right at the same time). The dynamic vibrotactile conditions (1 and 2) produced comparable braking reaction times that were faster than those in the static condition (3).

In a second experiment, Ahtamad et al. (2016) compared the expanding vibrotactile condition against a linear motion condition (vibrotactile stimulation to the hands followed by the shoulders). Emergency braking reaction times were faster in the linear motion condition (approximately 585 ms vs 640 ms) because drivers found it easier to interpret the warning signals in that condition.

In sum, the various auditory and vibrotactile warning signals discussed above typically reduce braking reaction times by approximately 40 ms. That sounds modest. However, it can easily be the difference between colliding with the car in front or avoiding it and so could potentially save many lives. At present, however, we lack a theoretical framework within which to understand precisely *why* some warning signals are more effective than others.

above and the other below fixation) and two auditory stimuli (low- and high-pitch). When the visual and auditory stimuli were *congruent* (e.g., visual stimulus above fixation and auditory stimulus high-pitch), the temporal ventriloquism effect was found. However, this effect was eliminated when the visual and auditory stimuli were *incongruent*, which prevented binding of information across the two senses.

COMPANION @ WEBSITE

Case study:
Multi-tasking efficiency

Overall evaluation

What are the limitations of research on cross-modal effects? First, as just mentioned, our theoretical understanding lags behind the accumulation of empirical findings. Second, much research has involved complex artificial tasks far removed from naturalistic conditions. Third, individual differences have generally been ignored. However, individual differences (e.g., preference for auditory or visual stimuli) influence cross-modal effects (van Atteveldt et al., 2014).

DIVIDED ATTENTION: DUAL-TASK PERFORMANCE

In this section, we consider factors influencing how well we can perform two tasks at the same time. In our hectic 24/7 lives, we increasingly try to do two things at once (**multi-tasking**) (e.g., sending text messages while walking down the street). More specifically, multi-tasking "refers to the ability to co-ordinate the completion of several tasks to achieve an overall goal" (MacPherson, 2018, p. 314). It can involve performing two tasks at the same time or switching between two tasks. There is controversy as to whether massive amounts of multi-tasking have beneficial or detrimental effects on attention and cognitive control (see Box).

What determines how well we can perform two tasks at once? *Similarity* (e.g., in terms of modality) is one important factor. Treisman and Davies (1973) found two monitoring tasks interfered with each other much more when the stimuli on both tasks were in the same modality (visual or auditory).

Two tasks can also be similar in response modality. McLeod (1977) had participants perform a continuous tracking task with manual responding together with a tone-identification task. Some participants responded vocally to the tones whereas others responded with the hand not involved in tracking. Tracking performance was worse with high response similarity (manual responses on both tasks) than with low response similarity.

Practice is the most important factor determining how well two tasks can be performed together. The saying "Practice makes perfect" was apparently supported by Spelke et al. (1976). Two students (Diane and John) received 5 hours of training a week for 4 months on various tasks. Their first task involved reading short stories for comprehension while writing down words from dictation, which they initially found very hard. After 6 weeks of training, however, they could read as rapidly and with as much comprehension when writing to dictation as when only reading. With further training, Diane and John learned to write down the names of the categories to which the dictated words belonged while maintaining normal reading speed and comprehension.

Spelke et al.'s (1976) findings are hard to interpret for various reasons. First, Spelke et al. focused on *accuracy* measures, which are typically less sensitive to dual-task interference than *speed* measures. Second, Diane and John's attentional focus was relatively uncontrolled, and so they may have *alternated* attention between tasks rather than attending to both at the same time. More controlled research on the effects of practice on dual-task performance is discussed later.

IN THE REAL WORLD: MULTI-TASKING 🌐

What are the effects of frequent multi-tasking in our everyday lives? Two main answers have been proposed. First, heavy multi-tasking may *impair* cognitive control because it leads individuals to allocate their attentional resources too widely. This is the scattered attention hypothesis (van der Schuur et al., 2015).

Second, heavy multi-tasking may *enhance* some control processes (e.g., task switching) because of prolonged practice in processing multiple streams of information. This is the trained attention hypothesis (van der Schuur et al., 2015). The relevant evidence is very inconsistent – "positive, negative, and null effects have all been reported" (Uncapher & Wagner, 2018, p. 9894).

Ophir et al. (2009) used a questionnaire (the Media Multitasking Index) to identify levels of multi-tasking. Heavy multi-taskers were more distractible. In a review, van der Schuur et al. (2015) found findings supported the scattered attention hypothesis (e.g., heavy multi-taskers had impaired sustained attention).

Moisala et al. (2016) found heavy multi-taskers were more adversely affected than light multi-taskers by distracting stimuli while performing speech–listening and reading tasks. During distraction, the heavy multi-taskers had greater activity than the light multi-taskers in the right pre-frontal cortex (associated with attentional control). This suggests heavy multi-taskers have greater problems than previously believed – their performance is impaired even though they try harder to exert top-down attentional control.

Uncapher and Wagner (2018) found in a review that most research indicated negative effects of heavy multi-tasking on tasks involving working memory, long-term memory, sustained attention and relational reasoning. These negative effects are likely to be due to attentional lapses. Of relevance, there are several studies where media multi-tasking was positively associated with self-reported everyday attentional failures. In addition, heavy multi-taskers often report high impulsivity – such individuals often make rapid decisions based on very limited evidence.

Most studies have only found an *association* between media multi-tasking and measures of attention and performance. This makes it hard to establish causality – it is possible individuals with certain patterns of attention choose to engage in extensive multi-tasking. Evidence suggesting that media multi-tasking can cause attention problems was reported by Baumgartner et al. (2018). They found that high media multi-tasking at one point in time predicted attention problems several months later.

Serial vs parallel processing

When individuals perform two tasks together, they might use serial or parallel processing. Serial processing involves switching attention backwards and forwards between two tasks with only one task being processed at any given moment. In contrast, parallel processing involves processing both tasks at the same time.

There has been much theoretical controversy on the issue of serial vs parallel processing in dual-task conditions (Koch et al., 2018). Of importance, processing can be mostly parallel or mostly serial. Lehle et al. (2009) trained participants to use serial or parallel processing when performing two tasks together. Those using serial processing performed better. However, they found the tasks more effortful because they had to inhibit processing of one task while performing the other one.

Lehle and Hübner (2009) also instructed participants to perform two tasks together in a serial or parallel fashion. Those using parallel processing

performed much worse. Fischer and Plessow (2015) reviewed dual-task research and concluded: "While serial task processing appears to be the most efficient [dual-task] processing strategy, participants are able to adopt parallel processing. Moreover, parallel processing can even outperform serial processing under certain conditions" (p. 8).

Brüning and Manzey (2018) confirmed serial processing is not always more efficient than parallel processing. Participants performed many alternate trials on two different tasks but could see the stimulus for the next trial ahead of time. Participants engaging in parallel processing (processing the stimulus for trial *n*+1 during trial *n*) performed better than those using only serial processing (*not* processing the trial *n*+1 stimulus ahead of time). Parallel processing reduced the costs incurred when task switching. Individuals high in working memory capacity (see Glossary) were more likely to use parallel processing, perhaps because of their superior attentional control.

IN THE REAL WORLD: CAN WE THINK AND DRIVE?

Car driving is the riskiest activity engaged in by tens of millions of adults. Over 50 countries have laws restricting the use of mobile or cell phones by drivers to increase car safety. Are such restrictions necessary? The short answer is "Yes" – drivers using a mobile phone are several times more likely to be involved in a car accident (Nurullah, 2015). This is so even though drivers try to reduce the risks by driving slightly more slowly (reducing speed by 5–6 mph) than usual shortly after initiating a mobile-phone call (Farmer et al., 2015).

Caird et al. (2008) in a review of studies using simulated driving tasks reported that reaction times to events (e.g., onset of brake lights on the car in front) increased by 250 ms with mobile-phone use and were greater when drivers were talking rather than listening. This 250 ms increase in reaction time translates into travelling an extra 18 feet (5.5 metres) before stopping for a motorist doing 50 mph (80 kph). This could be the difference between stopping just short of a child or killing that child.

Strayer and Drews (2007) studied the above slowing effect using event-related potentials while drivers responded rapidly to the onset of brake lights on the car in front. The magnitude of the P300 (a positive wave associated with attention) was reduced by 50% in mobile-phone users.

Strayer et al. (2011) considered a real-life driving situation. Drivers were observed to see whether they obeyed a law requiring them to stop at a road junction. Of drivers not using a mobile phone, 79% obeyed the law compared to only 25% of mobile-phone users.

Theoretical considerations

Why do so many drivers endanger people's lives by using mobile phones? Most believe they can drive safely while using a mobile phone whereas other drivers cannot (Sanbonmatsu et al., 2016b). Their misplaced confidence depends on limited monitoring of their driving performance: drivers using a mobile phone make more driving errors but do not remember making more errors (Sanbonmatsu et al., 2016a).

Why does mobile-phone use impair driving performance? Strayer and Fisher (2016) in their SPIDER model identified five cognitive processes that are adversely affected when drivers' attention is diverted from driving (e.g., by mobile-phone use):

(1) There is less effective visual scanning of the environment for potential threats. Distracted drivers are more inclined to focus attention on the centre of the road and less inclined to scan objects in the periphery and their side mirrors (Strayer & Fisher, 2016).

(2) The ability to predict *where* threats might occur is impaired. Distracted drivers are much less likely to make anticipatory glances towards the location of a potential hazard (e.g., obstructed view of a pedestrian crossing) (Taylor et al., 2015).

(3) There is reduced ability to identify visible threats, a phenomenon known as inattentional blindness (see Glossary; and Chapter 4). In a study by Strayer and Drews (2007), 30 objects (e.g., pedestrians; advertising hoardings) were clearly visible to drivers. However, those using a mobile phone subsequently recognised far fewer objects they had fixated than those not using a mobile phone (under 25% vs 50%, respectively).

(4) It is harder to decide what action is necessary in a threatening situation. Cooper et al. (2009) found drivers were 11% more likely to make unsafe lane changes when using a mobile phone.

(5) It becomes harder to execute the appropriate action. Reaction times are slowed (Caird et al., 2008, discussed above, p. 214).

The SPIDER model is oversimplified in several ways. First, various different activities are associated with mobile-phone use. Simmons et al. (2016) found in a meta-analytic review that the risk of safety-critical events was increased by activities requiring drivers to take their eyes off the road (e.g., locating a phone; dialling; texting). However, talking on a mobile phone did *not* increase risk.

Second, driving-irrelevant cognitive activities do not *always* impair all aspects of driving performance. Engstrom et al. (2017, p. 734) proposed their cognitive control hypothesis: "*Cognitive load selectively impairs driving sub-tasks that rely on cognitive control but leaves automatic performance unaffected.*" For example, driving-irrelevant activities involving cognitive load (e.g., mobile-phone use) typically have no adverse effect on well-practised driving skills, such as lane keeping and braking when getting close to the vehicle in front (Engstrom et al., 2017).

Third, individuals using mobile phones while driving are *unrepresentative* of drivers in general (e.g., they tend to be relatively young and to engage in more risk-taking activities: Precht et al., 2017). Thus, we must consider individual differences in personality and risk taking when interpreting accidents associated with mobile-phone use.

Fourth, the SPIDER model implies that performance cannot be *improved* by adding a secondary task. However, driving performance in monotonous conditions is sometimes *better* when drivers listen to the radio at the same time (see Engstrom et al., 2017). Listening to the radio can reduce the mind-wandering that occurs when someone drives in monotonous conditions. Drivers indicating their immediate thoughts during their daily commute reported mind-wandering 63% of the time and active focus on driving only 15%–20% of the time (Burdett et al., 2018).
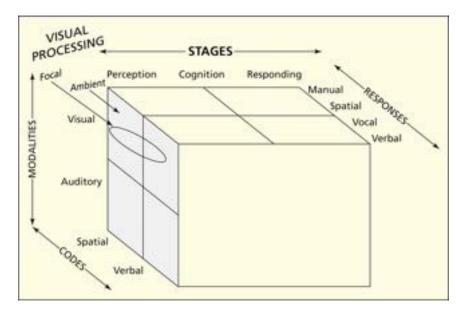
Multiple resource theory

Wickens (1984, 2008) argued in his multiple resource model that the processing system consists of several independent processing resources or mechanisms. The model includes four major dimensions (see Figure 5.16):

(1) *Processing stages*: there are successive stages of perception, cognition (e.g., working memory) and responding.

(2) *Processing codes*: perception, cognition and responding can use spatial and/or verbal codes; action can involve speech (vocal verbal) or manual/spatial responses.

(3) *Modalities*: perception can involve visual and/or auditory resources.
(4) *Visual channels*: visual processing can be focal (high acuity) or ambient
     (peripheral).

Here is the model's crucial prediction: "*To the extent that two tasks use dif-
ferent levels along each of the three dimensions* [excluding (4) above]*, time-
sharing* [dual-task performance] *will be better*" (Wickens, 2008, p. 450). Thus,
tasks requiring different resources can be performed together more success-
fully than those requiring the same resources. Wickens's approach bears
some resemblance to Baddeley's (e.g., 2012) working memory model (see
Chapter 6). According to that model, two tasks can be performed together
successfully provided they use *different* components or processing resources.

Findings

Research discussed earlier (Treisman & Davies, 1973; McLeod, 1977)
showing the negative effects of stimulus and response similarity on per-
formance are entirely consistent with the theory. Lu et al. (2013) reviewed
research where an ongoing visual-motor task (e.g., car driving) was
performed together with an interrupting task in the visual, auditory or
tactile (touch) modality. As predicted, non-visual interrupting tasks (espe-
cially those in the tactile modality) were processed more effectively than
visual ones and there were no adverse effects on the visual-motor task.

    According to the model, there should be only limited dual-task inter-
ference between two visual tasks if one requires focal or foveal vision,
whereas the other requires ambient or peripheral vision. Tsang and Chan
(2018) obtained support for this prediction in a study in which participants
tracked a moving target in focal vision while responding to a spatial task
in ambient or peripheral vision.

    Dual-task performance is often more impaired than predicted by
the theory. For example, consider a study by Robbins et al. (1996; see

Chapter 6). The main task was selecting chess moves and we will focus on the condition where the task performed at the same time was generating random letters. These two tasks involve different processing codes (spatial vs verbal, respectively) and they also involve different response types (manual vs vocal, respectively). Nevertheless, generating random letters caused substantial interference on the chess task.

Evaluation

The main assumptions of the theory have largely been supported by the experimental evidence. In other words, dual-task performance is generally less impaired when two tasks differ with respect to modalities, processing codes or visual channels than when they do not.

What are the model's limitations?

(1) Successful dual-task performance often requires higher-level processes of coordinating and organising the demands of the two tasks (see later section on cognitive neuroscience, pp. 220–222). However, these processes are de-emphasised within the theory.

(2) The theory's assumption there is a sequence of processing stages (perception; cognition; responding) is too rigid given the flexible nature of much dual-task processing (Koch et al., 2018). The numerous forms of cognitive processing intervening between perception and responding are not discussed in detail.

(3) It is implied within the theory that negative or interfering effects of performing two tasks together would be *constantly* present. However, Steinborn and Huestegge (2017) found dual-task conditions led only to *occasional* performance breakdown due to attention failures.

Threaded cognition

Salvucci and Taatgen (2008, 2011) proposed a model of threaded cognition in which streams of thought are represented as threads of processing. For example, processing two tasks might involve two separate threads. The central theoretical assumptions are as follows:

> Multiple threads or goals can be active at the same time, and as long as there is no overlap in the cognitive resources needed by these threads, there is no multi-tasking interference. When threads require the same resource at the same time, one thread must wait and its performance will be adversely affected.
>
> (Salvucci & Taatgen, 2011, p. 228)

This is because all resources have limited capacity.

Taatgen (2011) discussed the threaded cognition model (see Figure 5.17). Several cognitive resources can be the source of competition between two tasks. These include visual perception, declarative memory, task control and focal working memory or problem state. Nijboer et al. (2016a) discussed similarities between this model and Baddeley's working memory model (see Chapter 6). Three components of the model relate to
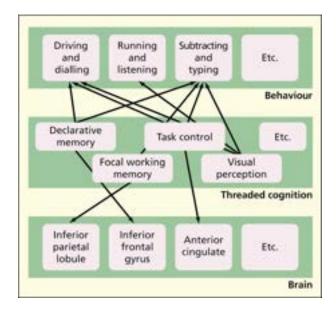
Figure 5.17
Threaded cognition theory. We possess several cognitive resources (e.g., declarative memory, task control, visual perception). These resources can be used in parallel but each resource can only work on one task at a time. Our ability to perform two tasks at the same time (e.g., driving and dialling, subtraction and typing) depends on the precise ways in which cognitive resources need to be used. The theory also identifies some of the brain areas associated with cognitive resources.
From Taatgen (2011). With permission of the author.

working memory: (1) problem state (attentional focus); (2) declarative memory (activated short-term memory); and (3) subvocal rehearsal (resembling the phonological loop; see Chapter 6).

Each thread or task controls resources in a greedy, polite way – threads claim resources greedily when required but release them politely when no longer needed. These aspects of the model lead to one of its most original assumptions – *several* goals (each associated with a given thread) can be active simultaneously.

The model resembles Wickens's multiple resource model: both models assume there are several independent processing resources. However, only the threaded cognition model led to a computational model making specific predictions. In addition, the threaded cognition model identifies the brain areas associated with each processing resource (see Figure 5.17).

Findings

According to the model, any given cognitive resource (e.g., visual perception; focal working memory) can be used by only *one* process at any given time. Nijboer et al. (2013) tested this assumption using multi-column subtraction as the *primary* task with participants responding using a keypad. Easy and hard conditions differed in whether digits were carried over ("borrowed") from one column to the next:

(1: easy)	(2: hard)
336789495	3649772514
–224578381	–1852983463

The model predicts focal working memory is required only in the hard condition. Subtraction was combined with a *secondary* task: a tracking task involving visual and manual resources or a tone-counting task involving working memory.

Nijboer et al. (2013) predicted performance on the easy subtraction task should be worse when combined with the tracking task because both compete for visual and manual resources. In contrast, performance on the hard subtraction task should be worse when combined with the tone-counting task because there are large disruptive effects when two tasks compete for working memory resources. The findings were as predicted.

Borst et al. (2013) found there was far less impairment of hard subtraction performance by a secondary task requiring working memory when

participants saw a visual sign explicitly indicating that "borrowing" was needed. This supports the model's assumption that dual-task performance can be enhanced by appropriate environmental support.

According to the threaded cognition model, we often cope with the demands of combining two tasks by switching *flexibly* between them to maximise performance. Support was reported by Farmer et al. (2018). Participants performed a typing task and a tracking task at the same time. The relative value of the two tasks was varied by manipulating the number of points lost for poor tracking performance. Participants rapidly learned to adjust their strategies over time to increase the overall number of points they gained.

Katidioti and Taatgen (2014) found task switching is not always optimal. Participants performed two tasks together: (1) an email task in which information needed to be looked up; (2) chat messages containing questions to be answered. When there was a delay on the email task, most participants switched to the chat task. This happened even when this was suboptimal because it caused participants to forget information in the email task.

How can we explain the above findings? According to Katidioti and Taatgen (2014, p. 734), "The results . . . agree with threaded cognition's 'greedy' theory . . . which states that people will switch to a task that is waiting as soon as the resources for it are available." Huijser et al. (2018) obtained further evidence of "greediness". When there were brief blank periods during the performance of a cognitively demanding task, participants often had task-irrelevant thoughts (e.g., mind-wandering) even though these thoughts impaired task performance.

Katidioti and Taatgen (2014) also discovered substantial individual differences in task switching – some participants never switched to the chat task when delays occurred on the email task. Such individual differences cannot be explained by the theory.

As mentioned earlier, a recent version of threaded cognition theory discussed by Nijboer et al. (2016a) identifies three components of working (i.e., problem state or focus of attention; declarative memory or activated short-term memory; and subvocal rehearsal). Nijboer et al. had participants perform two working memory tasks at the same time; these tasks varied in the extent to which they required the same working memory components. They obtained measures of performance and also used neuroimaging under dual-task and single-task conditions.

What did Nijboer et al. (2016a) find? First, dual-task interference could be predicted from the extent to which the two tasks involved the same working memory components. Second, dual-task interference could also be predicted from the extent of overlap in brain activation of the two tasks in single-task conditions. In sum, dual-task interference depended on competition for specific resources (i.e., working memory components) rather than general resources (e.g., central executive).

Evaluation

The model has proved successful in various ways. First, several important cognitive resources have been identified. Second, the model identifies brain

areas associated with various cognitive resources. This has led to computational modelling testing the model's predictions using neuroimaging and behavioural findings. Thus, the model accounts for dual-task performance without assuming the existence of a central executive or other executive process (often vaguely defined in other theories). Fourth, the theory predicts factors determining switching between two tasks being performed together. Fifth, individuals often have fewer problems performing two simultaneous tasks than generally assumed.

What are the model's limitations? First, it predicts that "Practising two tasks concurrently [together] results in the same performance as performing the two tasks independently" (Salvucci & Taatgen, 2008, p. 127). This de-emphasises the importance of processes coordinating and managing two tasks performed together (see next section). Second, excluding processes resembling Baddeley's central executive is controversial and may well prove inadvisable. Third, most tests of the model have involved the simultaneous performance of two relatively simple tasks and its applicability to more complex tasks remains unclear. Fourth, the theory does not provide a full explanation for individual differences in the extent of task switching (e.g., Katidioti & Taatgen, 2014).
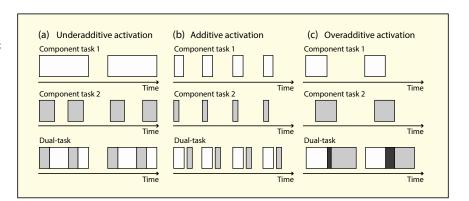
Cognitive neuroscience

The cognitive neuroscience approach is increasingly used to test theoretical models and enhance our understanding of processes underlying dual-task performance. Its value is that neuroimaging provides "an additional data source for contrasting between alternative models" (Palmeri et al., 2017, p. 61). More generally, behavioural findings indicate the extent to which dual-task conditions impair task performance but are often relatively uninformative about the precise reasons for such impairment.

Suppose we compare patterns of brain activation while participants perform tasks *x* and *y* singly or together. Three basic patterns are shown in Figure 5.18:

(1) *Underadditive activation*: reduced activation in one or more brain areas in the dual-task condition occurs because of resource competition between the tasks.

Figure 5.18
(a) Underadditive activation; (b) additive activation; (c) overadditive activation. White indicates task 1 activation; grey indicates task 2 activation; and black indicates activation only present in dual-task conditions.

From Nijboer et al., 2014. Reprinted with permission of Elsevier.

(2) *Additive activation*: brain activation in the dual-task condition is simply the *sum* of the two single-task activations because access to resources is integrated efficiently between the two tasks.

(3) *Overadditive activation*: brain activation in one or more brain areas is present in the dual-task condition but not the single-task conditions. This occurs when dual-task conditions require executive processes that are absent (or less important) with single tasks. These executive processes include the coordination of task demands, attentional control and dual-task management generally. We would expect such executive processes to be associated mostly with activation in prefrontal cortex.

> **KEY TERM** 🔑
>
> **Underadditivity**
> The finding that brain activation when tasks A and B are performed at the same time is less than the sum of the brain activation when tasks A and B are performed separately.

Findings

We start with an example of underadditive activation. Just et al. (2001) used two very different tasks (auditory sentence comprehension and mental rotation of 3-D figures) performed together or singly. Performance on both tasks was impaired under dual-task conditions compared to single-task conditions. Under dual-task conditions, brain activation in language processing areas decreased by 53% and reduced by 29% in areas associated with mental rotation. These findings suggest fewer task-relevant processing resources were available when both tasks were performed together.

Schweizer et al. (2013) also reported underadditivity. Participants performed a driving task on its own or with a distracting secondary task (answering spoken questions). Driving performance was unaffected by the secondary task. However, driving with distraction *reduced* activation in posterior brain areas associated with spatial and visual processing (underadditivity). It also produced *increased* activation in the prefrontal cortex (overadditivity; see Figure 5.19) probably because driving with distraction requires increased attentional or cognitive control within the prefrontal cortex.
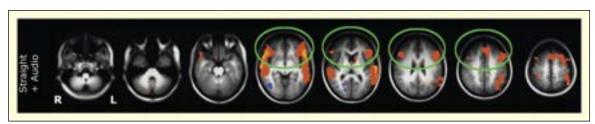
Dual-task performance is often associated with overadditivity due to increased activity within the prefrontal cortex (especially the lateral prefrontal cortex) during dual-task performance (see Strobach et al., 2018, for a review). However, most such findings do not show that this increased prefrontal activation is actually *required* for dual-task performance.

More direct evidence that prefrontal areas associated with attentional or cognitive control are *causally* involved in enhancing dual-task performance was reported by Filmer et al. (2017) and Strobach et al. (2018). Filmer et al. (2017) studied the effects of transcranial direct current stimulation (tDCS; see Glossary) applied to areas of the prefrontal cortex associated with cognitive control. Anodal tDCS during training enhanced cognitive control and subsequent dual-task performance.

Figure 5.19
Effects of an audio distraction task on brain activity associated with a straight driving task. There were significant increases in activation within the ventrolateral prefrontal cortex and the auditory cortex (in orange). There was decreased activation in occipital-visual areas (in blue).
From Schweizer et al. (2013).

Strobach et al. (2018) reported similar findings. Anodal tDCS applied to the lateral prefrontal cortex led to *enhanced* dual-task performance. In another condition, cathodal tDCS to the same area of the prefrontal cortex *impaired* dual-task performance. These findings were as predicted given that anodal and cathodal tDCS often have opposite effects on performance. These findings indicate that the lateral prefrontal cortex causally influences dual-task performance.

Additional evidence of the importance of the lateral prefrontal cortex was reported by Wen et al. (2018). Individuals with high connectivity (connectedness) within that brain area showed superior dual-task performance to those with low connectivity.

Finally, *patterns* of brain activation can help to explain practice effects on dual-task performance. Garner and Dux (2015) found much fronto-parietal activation (associated with cognitive control) when two tasks were performed singly or together. Extensive training greatly reduced dual-task interference and also produced increasing *differentiation* in the pattern of fronto-parietal activation associated with the two tasks. Participants showing the greatest reduction in dual-task interference tended to have the greatest increase in differentiation. Thus, using practice to increase differences in processing and associated brain processing between tasks can be very effective.

Evaluation

Brain activity in dual-task conditions often differs from the sum of brain activity of the same two tasks performed singly. Dual-task activity can exhibit underadditivity or overadditivity. The findings are theoretically important because they indicate performance of dual tasks can involve much more cognitive control and other processes than single tasks. Garner and Dux's (2015) findings demonstrate that enhanced dual-task performance with practice can depend on increased differentiation between the two tasks with respect to processing and brain activation.

What are the limitations of the cognitive neuroscience approach? First, increased (or decreased) activity in a given brain area in dual-task conditions is not necessarily very informative. For example, Dux et al. (2009) found dual-task performance improved over time because practice increased the speed of information processing in the prefrontal cortex rather than because it changed activation within that region. Second, it is often unclear whether patterns of brain activation are *directly* relevant to task processing rather than reflecting non-task processing.

Third, findings in this area are rather inconsistent (Strobach et al., 2018) and we lack a comprehensive theory to account for these inconsistencies. Plausible reasons for these apparent inconsistencies are the great variety of task combinations used in dual-task studies and individual differences in task proficiency among participants (Watanabe & Funahashi, 2018).

Psychological refractory period: cognitive bottleneck?

Much of the research discussed so far was limited because the task combinations used made it hard to assess in detail the processes used by participants. For example, the data collected were often insufficient to indicate

the frequency with which participants switched their attentional focus from one task to the other. This led researchers to use much simpler tasks so that they had "better experimental control over the timing of the component task processes" (Koch et al., 2018, p. 575).

The dominant paradigm in recent research is as follows. There are two stimuli (e.g., two lights) and two responses (e.g., button presses), one associated with each stimulus. Participants respond to each stimulus as rapidly as possible. When the two stimuli are presented at the same time (dual-task condition), performance is typically worse on both tasks than when each task is presented separately (single-task conditions).

When the second stimulus is presented shortly after the first, there is typically a marked slowing of the response to the second stimulus: the **psychological refractory period (PRP) effect**. This effect is robust – Ruthruff et al. (2009) obtained a large PRP effect even when participants were given strong incentives to eliminate it.

The PRP effect has direct real-world relevance. Hibberd et al. (2013) studied the effects of a simple in-vehicle task on braking performance when the vehicle in front braked and slowed down. There was a classic PRP effect – braking time was slowest when the in-vehicle task was presented immediately before the vehicle in front braked.

How can we explain the PRP effect? It is often argued task performance involves three successive stages: (1) perceptual; (2) central response selection; and (3) response execution. According to the bottleneck model (e.g., Pashler, 1994),

> The response selection stage of the second task cannot begin until the response selection stage of the first task has finished, although the other stages . . . can proceed in parallel . . . according to this model, the PRP effect is a consequence of the waiting time of the second task because of a bottleneck at the response selection stage.
>
> (Mittelstädt & Miller, 2017, p. 89)

The bottleneck model explains several findings. For example, consider the effects of varying the time between the start of the first and second stimuli (**stimulus onset asynchrony (SOA)**). According to the model, processing on the first task should slow down second-task processing much more when the SOA is small than when it is larger. The predicted finding is generally obtained (Mittelstädt & Miller, 2017).

The bottleneck model remains the most influential explanation of the PRP effect (and other dual-task costs). However, resource models (e.g., Navon & Miller, 2002) are also influential. According to these models, limited processing capacity can be *shared* between two tasks so both are processed simultaneously. Of crucial importance, sharing is possible even during the response selection process. A consequence of sharing processing capacity across task is that each task is processed more slowly than if performed on its own.

Many findings can be explained by both models. However, resource models are more *flexible* than bottleneck models. Why is this? Resource models assume the division of processing resources between two tasks varies freely to promote efficient performance.

Another factor influencing the PRP effect is **crosstalk** (the two tasks interfere directly with each other). This mostly occurs when the stimuli and/or responses on the two tasks are similar. A classic example of crosstalk is when you try to rub your stomach in circles with one hand while patting your head with the other hand (try it!).

Finally, note that participants in most studies receive only modest amounts of practice in performing two tasks at the same time. As a consequence, the PRP effect may occur at least in part because participants receive insufficient practice to eliminate it.

Findings

According to the bottleneck model, we would expect to find a PRP effect even when easy tasks are used and/or participants receive prolonged practice. Contrary evidence was reported by Schumacher et al. (2001). They used two tasks: (1) say "one", "two" or "three" to low-, medium- and high-pitched tones, respectively; (2) press response keys corresponding to the position of a disc on a computer screen. These tasks were performed together for over 2,000 trials, by which time some participants performed them as well together as singly.

Strobach et al. (2013) conducted a study very similar to that of Schumacher et al. (2001). Participants took part in over 5,000 trials involving single-task or dual-task conditions. However, dual-task costs were *not* eliminated after extensive practice: dual-task costs on the auditory task reduced from 185 to 60 ms and those on the visual task from 83 to 20 ms (see Figure 5.20). How did dual-task practice benefit performance? Practice speeded up the central response selection stage in both tasks.

Why did the findings differ in the two studies discussed above? In both studies, participants were rewarded for fast responding on single-task and dual-task trials. However, the way the reward system was set up in the Schumacher et al. study may have led participants to exert more effort in dual-task than single-task trials. This potential bias was absent from the
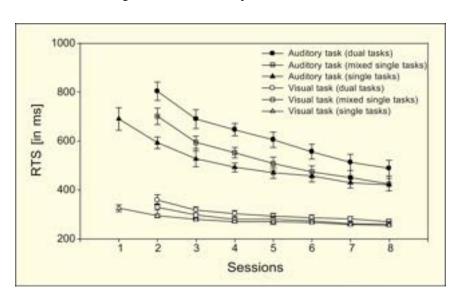
Figure 5.20
Reaction times for correct responses only over eight experimental sessions under dual-task (auditory *and* visual tasks) and single-task (auditory *or* visual task) conditions.

From Strobach et al. (2013). Reprinted with permission of Springer.

Strobach et al. study. This difference in reward structure could explain the greater dual-task costs in the Strobach et al. study.

Hesselmann et al. (2011) studied the PRP effect using event-related potentials. The slowing of responses on the second task was closely matched by slowing in the onset of the P300 (an ERP component reflecting response selection). However, there was *no* slowing of earlier ERP components reflecting perceptual processing. Thus, as predicted by the bottleneck model, the PRP effect depended on response selection rather than perceptual processes.

According to the resource model approach, individuals *choose* whether to use serial or parallel processing on PRP tasks. Miller et al. (2009) argued that serial processing generally leads to superior performance compared with parallel processing. However, parallel processing should theoretically be superior when the stimuli associated with the two tasks are mostly presented close in time. As predicted, there was a shift from predominantly serial processing towards parallel processing when that was the case.

Miller et al. (2009) used very simple tasks and it is likely parallel processing is most likely to be used with such tasks. Han and Marois (2013) used two tasks, one of which was relatively difficult. Participants used serial processing even when parallel processing was encouraged by financial rewards.

Finally, we consider the theoretically important **backward crosstalk effect**: "characteristics of Task 2 of 2 subsequently performed tasks influence Task 1 performance" (Janczyk et al., 2018, p. 261). Hommel (1998) obtained this effect. Participants responded to Task 1 by making a left or right key-press and to Task 2 by saying "left" or "right". Task 1 responses were faster when the two responses were compatible (e.g., press right key + say "right") than when they were incompatible (e.g., press right key + say "left"). Evidence for the backward crosstalk effect was also reported by Janczyk et al. (2018).

Why is the backward crosstalk effect theoretically important? It indicates that aspects of response selection processing on Task 2 occur *before* response selection processing on Task 1 has finished. This effect is incompatible with the bottleneck model, which assumes response selection on Task 1 is completed *prior* to any response selection on Task 2. In other words, this model assumes there is serial processing at the response selection stage. In contrast, the backward crosstalk effect is compatible with the resource model approach.

Summary and conclusions

The findings from most research on the psychological refractory period effect are consistent with the bottleneck model. As predicted, this effect is typically larger when the second task follows very soon after the first task. In addition, even prolonged practice rarely eliminates the psychological refractory period effect suggesting that central response selection processes typically occur serially.

The bottleneck model assumes processing is less *flexible* than is often the case. For example, the existence of the backward crosstalk effect is inconsistent with the bottleneck model but consistent with the resource

> **KEY TERM** 🔑
>
> **Backward crosstalk effect**
> Aspects of Task 2 influence response selection and performance speed on Task 1 in studies on the **psychological refractory period (PRP) effect**.

model approach. Fischer et al. (2018) also found evidence for much flexibility. There was less interference between the two tasks when financial rewards were offered because participants devoted more processing resources to protecting the first task from interference. However, the resource model approach has the disadvantage compared to the bottleneck model that its predictions are less precise, making it harder to submit to detailed empirical testing.

Finally, as Koch et al. (2017, p. 575) pointed out, the bottleneck model "can be applied (with huge success) mainly for conditions in which two tasks are performed strictly sequentially". This is often the case with research on the psychological refractory period effect but is much less applicable to more complex dual-task situations.

"AUTOMATIC" PROCESSING

We have seen in studies of divided attention that practice often causes a dramatic improvement in performance. This improvement has been explained by assuming some processes become *automatic* through prolonged practice. For example, the huge amount of practice we have had with reading words has led to the assumption that familiar words are read "automatically". Below we consider various definitions of "automaticity". We also consider different approaches to explaining the development of automatic processing.

Traditional approach: Shiffrin and Schneider (1977)

Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) distinguished between controlled and automatic processes:

- Controlled processes are of limited capacity, require attention and can be used flexibly in changing circumstances.
- Automatic processes suffer no capacity limitations, do not require attention and are very hard to modify once learned.

In Schneider and Shiffrin's (1977) research, participants memorised letters (the memory set) followed by a visual display containing letters. They then decided rapidly whether any item in the visual display was the same as any item in the memory set. The crucial manipulation was the type of mapping. With *consistent* mapping, only consonants were used as members of the memory set and only numbers were used as distractors in the visual display (or vice versa). Thus, a participant given only consonants to memorise would know any consonant detected in the visual display was in the memory set. With *varied* mapping, numbers and consonants were both used to form the memory set and to provide distractors in the visual display.

The mapping manipulation had dramatic effects (see Figure 5.21). The numbers of items in the memory set and visual display greatly affected decision speed *only* with varied mapping. According to Schneider and Shiffrin (1977), varied mapping involved *serial* comparisons between each item in the memory set and each item in the visual display until a match was achieved or every comparison had been made. In contrast, consistent
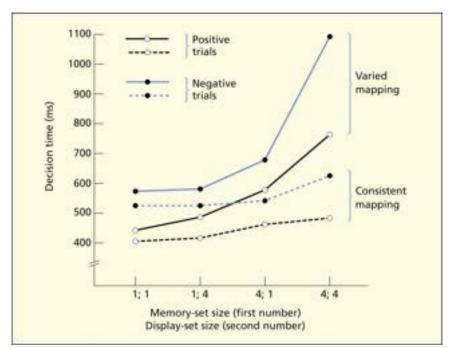
Figure 5.21
Response times on a decision task as a function of memory-set size, display-set size and consistent vs varied mapping.
Response times on a decision task as a function of memory-set size, display-set size and consistent vs varied mapping. Data from Shiffrin and Schneider (1977).

American Psychological Association.

mapping involved automatic processes operating independently and in *parallel*. These automatic processes have evolved through prolonged practice in distinguishing between letters and numbers.

In a second experiment, Shiffrin and Schneider (1977) used consistent mapping with the consonants B to L forming one set and Q to Z the other set. As before, items from only one set *always* formed the memory set with all the distractors in the visual display being selected from the other set. Performance improved greatly over 2,100 trials, reflecting increased automaticity. After that, there were 2,100 trials with the *reverse* consistent mapping (swapping over the memory and visual display sets). With this reversal, it took nearly 1,000 trials before performance recovered to its level at the start of the experiment!

Evidence that there may be limited (or no conscious awareness in the consistent mapping condition was reported by Jansma et al. (2001)). Increasing automaticity (indexed by increased performance speed) was accompanied by *reduced* activation in areas associated with conscious awareness (e.g., dorsolateral prefrontal cortex).

In sum, automatic processes function rapidly and in parallel but are inflexible (second part of the second experiment). Controlled processes are flexible and versatile but operate relatively slowly and in a serial fashion.

Limitations

What are the limitations with this approach? First, the distinction between automatic and controlled processes is oversimplified (discussed below). Second, Shiffrin and Schneider (1977) argued automatic processes operate in parallel and place no demands on attentional capacity and so decision speed should be unrelated to the number of items. However, decision

speed was slower when the memory set and visual display both contained several items (see Figure 5.21). Third, the theory is descriptive rather than explanatory – it does not explain how serial controlled processing turns into parallel automatic processing.

Definitions of automaticity

Shiffrin and Schneider (1977) assumed there is a clear-cut distinction between automatic and controlled processes. More specifically, automatic processes possess several features (e.g., inflexibility; very efficient because they have no capacity limitations; occurring in the absence of attention). In essence, it is assumed there is perfect *coherence* or consistency among the features (i.e., they are all found together).

Moors and De Houwer (2006) and Moors (2016) identified *four* key features associated with automaticity:

(1) unconscious: lack of conscious awareness of at least one of the following: "the input, the output, and the transition from one to the other" (Moors, 2016, p. 265);
(2) efficient: using very little attentional capacity;
(3) fast;
(4) goal-unrelated or goal-uncontrolled: at least one of the following is missing: "the goal is absent, the desired state does not occur, or the causal relation [between the goal and the occurrence of the desired state] is absent" (Moors, 2016, p. 265).

Why might these four features (or the similar ones identified by Shiffrin and Schneider (1977)) often be found together? Instance theory (Logan, 1988; Logan et al., 1999) provides an influential answer. It is assumed task practice leads to storage of information in long-term memory which facilitates subsequent performance on that task. In essence, "Automaticity is memory retrieval: performance is automatic when it is based on a single-step direct-access retrieval of past solutions from memory" (Logan, 1988, p. 493). For example, if you were given the problem "24 × 7 = ???" numerous times, you would retrieve the answer (168) "automatically" without performing any mathematical calculations.

Instance theory makes coherent sense of several characteristics of automaticity. Automatic processes are fast because they require only the retrieval of past solutions from long-term memory. They make few demands on attentional resources because the retrieval of heavily over-learned information is relatively effortless. Finally, there is no conscious awareness of automatic processes because no significant processes intervene between stimulus presentation and retrieval of the correct response.

In spite of its strengths, instance theory is limited (see Moors, 2016). First, the theory implies the key features of automaticity will typically all be found together. However, this is *not* the case (see below). Second, it is assumed practice leads to automatic retrieval of solutions with learners having no control over such retrieval. However, Wilkins and Rawson (2011) found evidence learners can exercise top-down control over retrieval: when the instructions emphasised accuracy, there was less evidence of retrieval

than when they emphasised speed. Thus, the use of retrieval after practice is *not* fully automatic.

Melnikoff and Bargh (2018) argued that the central problem with the traditional approach is that no research has shown the four features associated with "automaticity" occurring together. As they pointed out, "No attempt has been made to estimate the probability of a process being intentional given that is conscious versus unconscious, or the probability of a process being controllable given that it is efficient versus inefficient, and so forth" (p. 282).

Decompositional approach: Moors (2016)

Moors and De Houwer (2006) and Moors (2016) argued that previous theoretical approaches are greatly oversimplified. Instead, they favoured a decompositional approach. According to this approach, the features of automaticity are clearly separable and are by no means always found together: "It is dangerous to draw inferences about the presence or absence of one feature on the basis of the presence or absence of another" (Moors & de Houwer, 2006, p. 320).

Moors and De Houwer (2006) also argued there is no firm dividing line between automaticity and non-automaticity. The features are continuous rather than all-or-none (e.g., a process can be fairly fast or slow; it can be partially conscious). As a result, most processes involve a *blend* of automaticity and non-automaticity. This approach is rather imprecise because few processes are 100% automatic or non-automatic. However, we can make relative statements (e.g., process *X* is more/less automatic than process *Y*).

Moors (2016) claimed the relationships between factors such as goals, attention and consciousness are much more complex than claimed within traditional approaches to "automaticity". This led her to develop a new theoretical account (see Figure 5.22). A key assumption is that all information
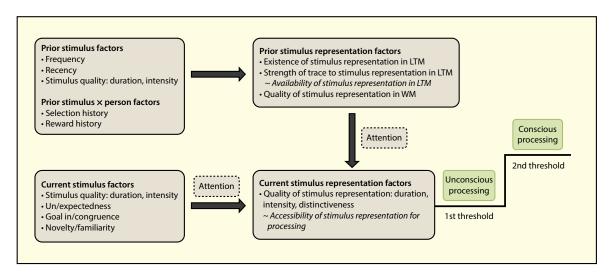

Figure 5.22

Factors that are hypothesised to influence representational quality within Moors' (2016) theoretical approach.

From Moors (2016).

processes require an input of sufficient representational quality (defined by the "intensity, duration, and distinctiveness of a representation", Moors, 2016, p. 273).

What factors determine representational quality?

(1) current stimulus factors, including the extent to which a stimulus is expected or unexpected, familiar or novel, and goal congruent or incongruent;

(2) prior stimulus factors (e.g., the frequency and recency with which the current stimulus has been encountered);

(3) prior stimulus representation factors based on relevant information stored within long-term memory;

(4) attention, which enhances or amplifies the impact of current stimulus factors and prior stimulus representation factors on the current stimulus representation.

According to this theoretical account, the above factors influence representational quality *additively* so that a high level of one factor can compensate for a low level of another factor. For example, selective attention or relevant information in long-term memory can compensate for brief stimulus presentations. The main impact of consciousness occurs later than for other factors (e.g., attention and goal congruence). More specifically, representational quality must reach the first threshold to permit unconscious processing but a more stringent second threshold to permit conscious processing.

Findings

According to Moors' (2016) theoretical framework, there is a *flexible* relationship between controlled and conscious processing. This contrasts with Schneider and Shiffrin's (1977) assumption that executive control is *always* associated with conscious processing.

Diao et al. (2016) reported findings consistent with Moors' prediction. They used a Go/No-Go task where participants made a simple response (Go trials) or withheld it (No-Go trials). High-value or low-value financial rewards were available for successful performance. Task stimuli were presented above or below the level of conscious awareness.

What did Diao et al. (2016) find? Performance was better on high-reward than low-reward trials even when task processing was unconscious. In addition, participants showed superior unconscious inhibitory control (assessed by event-related potentials) on high-reward trials. Thus, one feature of automaticity (unconscious processing) was present whereas another feature (goal-uncontrolled) was not.

Huber-Huber and Ansorge (2018) also reported problems for the traditional approach. Participants received target words indicating an upward or downward direction (e.g., *above*; *below*). Prior to the target word, a prime word also indicating an upward or downward direction was presented below the level of conscious awareness. Response times to the target words were slower when there was a conflict between the meanings of the prime and target words than when they were congruent in meaning. As in

the study by Diao et al. (2016), unconscious processing was combined with control, a combination that is inconsistent with the traditional approach.

Evaluation

The theoretical approach to automaticity proposed by Moors (2016) has several strengths. First, the assumption that various features associated with automaticity often correlate poorly with each other is clearly superior to the earlier notion that these features exhibit perfect coherence. Second, her assumption that processes vary in the extent to which they are "automatic" is much more realistic than the simplistic division of processes into automatic and non-automatic. Third, the approach is more comprehensive than previous ones because it considers more factors relevant to "automaticity".

What are the limitations with Moors' (2016) approach? First, numerous factors are assumed to influence representational quality (and thus the extent to which processes are automatic) (see Figure 5.22). It would thus require large-scale experimental research to assess the ways all these factors interact. Second, the approach provides only a partial explanation of the underlying mechanisms causing the various factors to influence representational quality.

Interactive exercise:
Definitions of attention

CHAPTER SUMMARY 📝

• **Focused auditory attention.** When two auditory messages are presented at the same time, there is less processing of the unattended than the attended message. Nevertheless, unattended messages often receive some semantic processing. The restricted processing of unattended messages may reflect a bottleneck at various stages of processing. However, theories assuming the existence of a bottleneck de-emphasise the flexibility of selective auditory attention. Attending to one voice among several (the cocktail party problem) is a challenging task. Human listeners use top-down and bottom-up processes to select one voice. Top-down processes include the use of various control processes (e.g., focused attention; inhibitory processes) and learning about structural consistencies present in the to-be-attended voice.

• **Focused visual attention.** Visual attention can resemble a spotlight or zoom lens. In addition, the phenomenon of split attention suggests visual attention can also resemble multiple spotlights. However, accounts based on spotlights or a zoom lens typically fail to specify the underlying mechanisms. Visual attention can be object-based, space-based or feature-based, and it is often object-based and space-based at the same time. Visual attention is flexible and is influenced by factors such as individual differences.

  According to Lavie's load theory, we are more susceptible to distraction when our current task involves low perceptual load and/or high cognitive load. There is much support for this theory.

However, the effects of perceptual and cognitive load are often not independent as predicted. In addition, it is hard to test the theory because the terms "perceptual load" and "cognitive load" are vague. There are stimulus-driven ventral attention and goal-directed dorsal attention networks involving different (but partially overlapping) brain networks. More research is required to establish how these two attentional systems interact. Additional brain networks (e.g., cingulo-opercular network; default mode network) relevant to attention have also been identified.

- **Disorders of visual attention.** Neglect occurs when damage to the ventral attention network in the right hemisphere impairs the functioning of the undamaged dorsal attention network. This impaired functioning of the dorsal attention network involves reduced activation and alertness within the left hemisphere. Extinction is due to biased competition for attention between the two hemispheres combined with reduced attentional capacity. More research is required to clarify differences among neglect patients in their specific processing deficits (e.g., the extent to which failures to detect left-field stimuli are due to impaired spatial working memory).

- **Visual search.** One problem with airport security checks is that there are numerous possible target objects. Another problem is the rarity of targets, which leads to excessive caution in reporting targets. According to feature integration theory, object features are processed in parallel and then combined by focused attention in visual search. This theory ignores our use of general scene knowledge in everyday life to focus visual search on areas of the scene most likely to contain the target object. It also exaggerates the prevalence of serial processing. Contemporary approaches emphasise the role of perception in visual search. Parallel processing is very common because much information is typically extracted from the peripheral visual field as well as from central or foveal vision. Problems in visual search occur when there is visual crowding in peripheral vision.

- **Cross-modal effects.** In the real world, we often coordinate information across sense modalities. In the ventriloquist effect, vision dominates sound because an object's location is typically indicated more precisely by vision. In the temporal ventriloquism effect, sound dominates vision because the auditory modality is typically more precise at discriminating temporal relations. Both effects depend on the assumption that visual and auditory stimuli come from the same object. Auditory or vibrotactile warning signals that are informative about the direction of danger and/or imminence of collision speed up drivers' braking times. We lack a theoretical framework within which to understand why some warning signals are more effective than others.

- **Divided attention: dual-task performance.** Individuals engaging in heavy multi-tasking show evidence of increased distractibility and impaired attentional control. A demanding secondary task (e.g., mobile-phone use) impairs aspects of driving performance requiring cognitive control but not well-practised driving skills (e.g., lane keeping). Multiple resource theory and threaded cognition theory both assume dual-task performance depends on several limited-capacity processing resources. This permits two tasks to be performed together successfully provided they use different processing resources. This general approach de-emphasises high-level executive processes (e.g., monitoring and coordinating two tasks).

  Some neuroimaging studies have found underadditivity in dual-task conditions (less activation than for the two tasks performed separately). This may indicate people have limited general processing resources. Other neuroimaging studies have found dual-task conditions can introduce new processing demands of task coordination associated with activation within the dorsolateral prefrontal cortex and cerebellum. It is often unclear whether patterns of brain activation are *directly* relevant to task processing.

  The psychological refractory period (PRP) effect can be explained by a processing bottleneck during response selection. This remains the most influential explanation. However, some evidence supports resource models claiming parallel processing of two tasks is often possible. Such models are more flexible than bottleneck models and they provide an explanation for interference effects from the second of two tasks on the first one.

- **"Automatic" processing.** Shiffrin and Schneider distinguished between slow, flexible controlled processes and fast, automatic ones. This distinction is greatly oversimplified. Other theorists have claimed automatic processes are unconscious, efficient, fast and goal-unrelated. However, these four processing features are not all-or-none and they often correlate poorly with each other. Thus, there is no sharp distinction between automatic and non-automatic processes. Moors' (2016) decompositional approach plausibly assumes that there is considerable flexibility in terms of the extent to which any given process is "automatic".

FURTHER READING

Chen, Y.-C. & Spence, C. (2017). Assessing the role of the "unity assumption" on multi-sensory integration: A review. *Frontiers in Psychology*, *8* (Article 445). Factors determining the extent to which stimuli from different sensory modalities are integrated are discussed.

Engstrom, J., Markkula, G., Victor, T. & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors*,

59, 734–764. Johan Engstrom and his colleagues review research on factors influencing driving performance and provide a new theoretical approach.

Hulleman, J. & Olivers, C.N.L. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, 40, 1–20. This review article indicates very clearly why theoretical accounts of visual search increasingly emphasise the role of fixations and visual perception. Several problems with previous attention-based theories of visual search are also discussed.

Karnath, H.-O. (2015). Spatial attention systems in spatial neglect. *Neuropsychologia*, *75*, 61–73. Hans-Otto Karnath discusses theoretical accounts of neglect emphasising the role of attentional systems.

Koch, I., Poljac, E., Müller, H. & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking – An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*, 557–583. Iring Koch and colleagues review dual-task and task-switching research with an emphasis on major theoretical perspectives.

McDermott, J.H. (2018). Audition. In J.T. Serences (ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. 2: Sensation, Perception, and Attention* (4th edn; pp. 63–120). New York: Wiley. Josh McDermott discusses theory and research focused on selective auditory attention in this comprehensive chapter.

Melnikoff, D.E. & Bargh, J.A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280–293. Research revealing limitations with traditional theoretical approaches to "automaticity" is discussed.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263–287. Agnes Moors provides an excellent critique of traditional views on "automaticity" and develops her own comprehensive theoretical account.

Nobre, A.C. (2018). Attention. In J.T. Serences (ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. 2: Sensation, Perception, and Attention* (4th edn; pp. 241–316). New York: Wiley. Anna (Kia) Nobre discusses the key role played by attention in numerous aspects of cognitive processing.

Taylor & Francis

Taylor & Francis Group

http://taylorandfrancis.com

Memory

PART II

How important is memory? Imagine if we were without it. We would not recognise anyone or anything as familiar. We would be unable to talk, read or write because we would remember nothing about language. We would have extremely limited personalities because we would have no recollection of the events of our own lives and therefore no sense of self. In sum, we would have the same lack of knowledge as a newborn baby.

Nairne et al. (2007) argued there were close links between memory and survival in our evolutionary history. Our ancestors prioritised information relevant to their survival (e.g., remembering the location of food or water; ways of securing a mate). Nairne et al. found memory for word lists was especially high when participants rated the words for their relevance to survival in a dangerous environment: the survival-processing effect. This effect has been replicated several times (Kazanas & Altarriba, 2015) and is stronger when participants imagine themselves alone in a dangerous environment rather than with a group of friends (Leding & Toglia, 2018). In sum, human memory may have evolved in part to promote survival.

We use memory for numerous purposes throughout every day of our lives. It allows us to keep track of conversations, to remember how to use a mobile phone, to write essays in examinations, to recognise other people's faces, to take part in conversations, to ride a bicycle, to carry out intentions and, perhaps, to play various sports. More generally, our *interactions* with others and with the environment depend crucially on having an effective memory system.

The wonders of human memory are discussed at length in Chapters 6–8. Chapter 6 deals mainly with key issues regarded as important from the early days of memory research. For example, we consider the distinction between short-term and long-term memory. The notion of short-term memory has been largely superseded by that of a working-memory system combining the functions of processing and short-term information storage. There is extensive coverage of working memory in Chapter 6.

Another topic discussed at length in Chapter 6 is learning. Long-term memory is generally enhanced when meaning is processed at the time of learning. Long-term memory is also better if much of the learning period

is spent practising retrieval. Evidence suggesting some learning is implicit (i.e., does not depend on conscious processes) is also discussed. Finally, we discuss forgetting. Why do we tend to forget information over time?

Chapter 7 is devoted to long-term memory. Our long-term memories include personal information, knowledge about language, much knowledge about psychology (hopefully!), knowledge about thousands of objects in the world around us, and information about how to perform various skills (e.g., riding a bicycle; playing the piano). The central issue addressed in Chapter 7 is how to account for this incredible richness. Several theorists have claimed there are several long-term memory systems. Others argue that there are numerous processes that are combined and recombined depending on the specific demands of any given memory task.

Memory is important in everyday life in ways de-emphasised historically. For example, autobiographical memory (discussed in Chapter 8) is of great significance to us. It gives us a coherent sense of ourselves and our personalities. The other topics considered in Chapter 8 are eyewitness testimony and prospective memory (memory for future intentions). Research into eyewitness testimony has revealed that eyewitness testimony is often much less accurate than generally assumed. This has implications for the legal system because hundreds of innocent individuals have been imprisoned solely on the basis of eyewitness testimony.

When we think about memory, we naturally focus on memory of the *past*. However, we also need to remember numerous *future* commitments (e.g., meeting a friend as arranged), and such remembering involves prospective memory. We will consider how we try to ensure we carry out our future intentions.

The study of human memory is fascinating, and substantial progress has been made. However, it is complex and depends on several factors. Four kinds of factors are especially important: events, participants, encoding and retrieval (Roediger, 2008). Events range from words and pictures to texts and life events. Participants vary in age, expertise, memory-specific disorders and so on. What happens at encoding varies as a function of task instructions, the immediate context and participants' strategies. Finally, memory performance at retrieval often varies considerably depending on the nature of the memory task (e.g., free recall; cued recall; recognition).

The take-home message is that memory findings are context-sensitive – they depend on *interactions* between the four factors. Thus, the effects of manipulating, say, what happens at encoding depend on the participants used, the events to be remembered and the conditions of retrieval. That explains why Roediger (2008) entitled his article, "Why the laws of memory vanished". How, then, do we make progress? As Baddeley (1978, p. 150) argued, "The most fruitful way to extend our understanding of human memory is not to search for broader generalisations and 'principles', but is rather to develop ways of separating out and analysing more deeply the complex underlying processes."

Learning, memory and forgetting

INTRODUCTION

This chapter (and the next two) focus on human memory. All three chapters deal with intact human memory, but Chapter 7 also considers amnesic patients in detail. Traditional laboratory-based research is the focus of this chapter and Chapter 7, with more naturalistic research being discussed in Chapter 8. There are important links among these different types of research. For example, many theoretical issues relevant to brain-damaged and healthy individuals can be tested in the laboratory or in the field.

Learning and memory involve several stages of processing. **Encoding** occurs during learning: it involves transforming presented information into a representation that can subsequently be stored. This is the first stage. As a result of encoding, information is stored within the memory system. Thus, *storage* is the second stage. The third stage is *retrieval*, which involves recovering information from the memory system. Forgetting (discussed later, see pp. 278–293) occurs when our attempts at retrieval are unsuccessful.

Several topics are discussed in this chapter. The basic structure of the chapter consists of three sections:

(1) The first section focuses mostly on short-term memory (a form of memory in which information is held for a brief period of time). This section has three topics (short-term vs long-term memory; working memory; and working memory: executive functions and individual differences). The emphasis here is on the early stages of processing (especially encoding).

(2) The second section focuses on learning and the processes occurring during the acquisition of information (i.e., encoding processes) leading to long-term memory. Learning can be explicit (occurring *with* conscious awareness of what has been learned) or implicit (occurring *without* conscious awareness of what has been learned). The first two topics in this section (levels of processing; learning through retrieval) focus on explicit learning whereas the third topic focuses on implicit learning.

> **KEY TERM**
>
> **Encoding**
> The process by which information contained in external stimuli is transformed into a representation that can be stored within the memory system.

(3) The third section consists of a single topic: forgetting from long-term memory. The emphasis differs from the other two sections in that the emphasis is on retrieval processes rather than encoding processes. More specifically, the focus is on the reasons responsible for the failures of retrieval.

SHORT-TERM VS LONG-TERM MEMORY

Many theorists distinguish between short-term and long-term memory. For example, there are enormous differences in capacity: only a few items can be held in short-term memory compared with essentially unlimited capacity in long-term memory. There are also massive differences in duration: a few seconds for short-term memory compared with up to several decades for long-term memory. The distinction between short-term and long-term memory stores was central to multi-store models. More recently, however, some theorists have proposed unitary-store models in which this distinction is much less clear-cut. Both types of models are discussed below.

Multi-store model

Atkinson and Shiffrin (1968) proposed an extremely influential multi-store model (see Figure 6.1):

- sensory stores, each modality-specific (i.e., limited to one sensory modality) and holding information very briefly;
- short-term store of very limited capacity;
- long-term store of essentially unlimited capacity holding information over very long periods of time.

According to the multi-store model, environmental stimulation is initially processed by the sensory stores. These stores are modality-specific (e.g., vision; hearing). Information is held very briefly in the sensory stores, with some being attended to and processed further within the short-term store.

Sensory stores

The visual store (iconic memory) holds visual information briefly. According to a recent estimate (Clarke & Mack, 2015), iconic memory for a natural scene lasts for at least 1,000 ms after stimulus offset. If you twirl
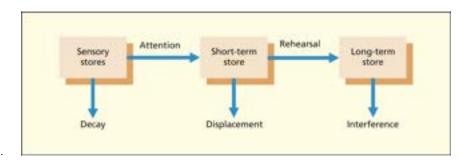

Figure 6.1
The multi-store model of memory as proposed by Atkinson and Shiffrin (1968).

a lighted object in a circle in the dark, you will see a circle of light because of the persistence of visual information in iconic memory. More generally, iconic memory increases the time for which visual information is accessible (e.g., when reading).

Atkinson and Shiffrin (1968) and many other theorists have assumed iconic memory is pre-attentive (not dependent on attention). However, Mack et al. (2016) obtained findings strongly suggesting that iconic memory does depend on attention. Participants had to report the letters in the centre of a visual array (iconic memory) or whether four circles presented close to the fixation point were the same colour. Performance on the iconic memory task was much worse when the probability of having to perform the iconic memory task was only 10% rather than 90%. This happened because there was much less attention to the letters in the former condition.

Echoic memory, the auditory equivalent of iconic memory, holds auditory information for a few seconds. Suppose someone asked you a question while your mind was elsewhere. Perhaps you replied "What did you say?", just before realising you did know what had been said. This "playback" facility depends on echoic memory. Ioannides et al. (2003) found the duration of echoic memory was longer in the left hemisphere than the right, probably because of the dominance of the left hemisphere in language processing.

There are sensory stores associated with all other senses (e.g., touch; taste). However, they are less important than iconic and echoic memory and have attracted much less research.

Short-term memory

Short-term memory has very limited capacity. Consider digit span: participants listen to a random digit series and then repeat back the digits immediately in the correct order. There are also letter and word spans. The maximum number of items recalled without error is typically about seven.

There are two reasons for rejecting seven items as the capacity of short-term memory. First, we must distinguish between items and chunks – "groups of items . . . collected together and treated as a single unit" (Mathy & Feldman, 2012, p. 346). For example, most individuals presented with the letter string *IBMCIAFBI* would treat it as three chunks rather than nine letters. Here is another example: you might find it hard to recall the following five words: is thing many-splendoured a love but easier to recall the same words presented as follows: love is a many-splendoured thing.

Simon (1974) showed the importance of chunking. Immediate serial recall was 22 words with 8-word sentences but only 7 with unrelated words. In contrast, the number of chunks recalled varied less: it was 3 with the sentences compared to 7 with the unrelated words. Second, estimates of short-term memory capacity are often inflated because participants' performance is influenced by rehearsal and long-term memory.

What influences chunking? As we have seen, it is strongly determined by information stored in long-term memory (e.g., IBM stands for International Business Machines). However, chunking also depends on people's abilities to identify patterns or regularities in the material presented for learning.

KEY TERMS

Echoic memory
A sensory store that holds auditory information for approximately 2–3 seconds.

Chunks
Stored units formed from integrating smaller pieces of information.

Interactive exercise:
Capacity of short-term memory

Interactive exercise:
Duration of short-term memory

For example, compare the digit sequences 2 3 4 5 6 and 2 4 6 3 5. It is much easier to chunk the former sequence as "all digits between 2 and 6". Chekaf et al. (2016) found participants' short-term memory was greatly enhanced by spontaneous detection of such patterns. When there were no patterns in the learning material, short-term memory was only three items.

A similar capacity limit was reported by Chen and Cowan (2009). When rehearsal was prevented by **articulatory suppression** (saying "the" repeatedly), only three chunks were recalled.

Within the multi-store model, it is assumed *all* items within short-term memory have equal importance. However, this is an oversimplification. Vergauwe and Langerock (2017) assessed speed of performance when participants were presented with four letters followed by a probe letter and decided whether the probe was the same as any of the original letters. Response to the probe was fastest when it corresponded to the letter currently being attended to (cues were used to manipulate which letter was the focus of attention at any given moment).

How is information lost from short-term memory? Several answers have been provided (Endress & Szabó, 2017). Atkinson and Shiffrin (1968) emphasised the importance of displacement – the capacity of short-term memory is very limited, and so new items often *displace* items currently in short-term memory. Another possibility is that information in short-term memory *decays* over time in the absence of rehearsal. A further possibility is *interference* which could come from items on previous trials and/or from information presented during the retention interval.

The experimental findings are variable. Berman et al. (2009) claimed interference is more important than decay. Short-term memory performance on any given trial was disrupted by words presented on the previous trial. Suppose this disruption effect occurred because words from the previous trial had not decayed sufficiently. If so, disruption would have been greatly reduced by *increasing* the inter-trial interval. In fact, increasing that interval had no effect. However, the disruption effect was largely eliminated when interference from previous trials was reduced.

Campoy (2012) pointed out Berman et al.'s (2009) research was limited because their experimental design did not allow them to observe any decay occurring within 3.3 seconds of item presentation. Campoy obtained strong decay effects at time intervals shorter than 3.3 seconds. Overall, the findings suggest decay occurs mostly at short retention intervals and interference at longer ones.

Strong evidence interference is important was reported by Endress and Potter (2014). They rapidly presented 5, 11 or 21 pictures of familiar objects. In their unique condition, no pictures were repeated over trials, whereas in their repeated condition, the same pictures were seen frequently over trials. Short-term memory was greater in the unique condition in which there was much less interference than in the repeated condition (see Figure 6.2).

In sum, most of the evidence indicates that interference is the most important factor causing forgetting from short-term memory, although decay may also play a part. There is little direct evidence that displacement (emphasised by Atkinson & Shiffrin, 1968) is the main factor causing forgetting. However, it is possible that interference causes items to be displaced from short-term memory (Endress & Szabó, 2017).